

## (TM)

\*\*\*\*\*

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	823	33.7	2449	29	HSU03109	Human aspartyl beta-hy	0.00e+00
2	821	33.6	2344	30	S83225	aspartyl(asparaginy)b	0.00e+00
3	214	8.8	2739	23	B0VABHY	Bos taurus aspartyl (a	1.25e-11
4	175	7.2	1777	23	CFU38414	Canis familiaris junct	6.81e-92
5	108	4.4	7218	25	166494	Sequence 14 from paten	4.21e-47
6	94	3.8	380	34	G23118	human STS WR-11767.	4.65e-38
7	84	3.4	86	25	AR003317	Sequence 39 from paten	1.04e-31
8	51	2.1	7218	25	166494	Sequence 14 from paten	8.25e-12
9	48	2.0	10772	21	AF012089	Drosophila melanogaste	3.99e-10
10	50	2.0	10772	21	AF012089	Drosophila melanogaste	3.03e-11
11	42	1.7	215	25	128278	Sequence 5 from paten	7.41e-07
12	36	1.2	965	25	AR042229	Sequence 22 from paten	7.41e-07
13	36	1.5	216021	31	HUAC004787	Homo sapiens Chromosom	9.39e-04

2 (bases 1 to 2739)  
 Friedman, P. A.  
 Direct Submission  
 Submitted (15-APR-1992) P. A. Friedman, Merck Sharp and Dohme  
 Research Laboratories, West Point, PA 19486 USA  
 Location/Qualifiers  
 1. 2739  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /tissue\_type="liver and brain"  
 189, 2453  
 /codon\_start=1  
 /product="aspartyl (asparaginyl) beta hydroxylase"  
 /protein\_id="AA03563.1"  
 /db\_xref="PID:g162694"  
 /db\_xref="GI:162694"  
 /translation="MAPRKNAGGGCGNSSSSSGSGPTGCTSGSSSPGARRETQKGL  
 KNGRKGSGSEFFWFWIALLGWVTSVAVVDFLDVDEVLAKAKDFRNLNLSVLQ  
 IKEDVYDAGDGDVDVDAKVLGLGKEKPAKPTYPPEADIMYPLWEDQVLESPGQ  
 IGEVYQVQSDJEDTVYSPGPELQPEEGPAELQDDHDFVFGDADRYEPMGTA  
 VHEEDTSYHIBETASLPASQWEDMMYEQENPDSSEPVVDDAERTYQETDDVTVR  
 YDHDADVNSNTILPEHPMAEEQOEVPETNKADPEKKGKKVKKPKLLNKFDF  
 KTKIAELDAEAKLRKRGKEEAVNAFEELVRKYPSQPGARYKQACEDDLAEKRRS  
 ILRAETIQAASLDPATDVLVLSLKRSDRQQLFLGMRGSLTLQKLVLQFLPDDT  
 ALKNDLGVGLYLIGDNSAKKVEEVLSTPNDGFAKVHYFLTKAQNIAIESIPYKL  
 EGIESGDDTDGREFHLGDAMQRVGNKEAYRWYELHGLGHGFASWORSYLRNVOGL  
 KAQWPTKETGYTELVKSLERNKLIIDEGLAAMDRTHGLFLPEDENLREKGDGFL  
 TLWQGRKNENACKGAPKTCSLLDKPEPTTGCRRGOIKYSIMHPGTHVMPHTGPTNR  
 LRLHGLVLPKSGCKTRCANETRWEEGKVLFDSDSEFEHVQDAASFRLLIFIVDYNH  
 PELTHQRKSLPAI"

BASE COUNT 801 a 636 c 749 g 553 t  
 ORIGIN

	Query Match	8.88;	Score 214;	DB 23;	Length 2739;
	Best Local Similarity	76.13;	Pred. No. 1.25e-118;		
	Matches 378;	Conservative 0;	Mismatches 110;	Indels 9;	Gaps 6;
Db	483	GAGTGACTTCAAGGAAAC	TTGGAATCTATGATGCTGATGGTGATGGAGATTTTGTGATGTG	542	
Qy	136	GAGGAAGTTCTAGAAAC	TAGGAATCTATGATGCTGATGGTGATGGATTTTGTGATGTG	195	
Db	543	GATGATGCCAAGTTTATT	TAGGCTTTAAAGAAAACCTGCTCCAAAGCCACAGTGCCA	602	
Qy	196	GATGATGCCAAGTTTATT	TAGGACTTTAAAGAGAGATCTACTTCAGAGCCAGCAGTCCCG	255	
Db	503	CCAGAAGAGCCAGACATG	TACCTTGGCTGGAGATCAGGTTCTTGGAGAGCCCGCAGGACGT	662	
Qy	256	CCAGAAGAGGCTGAGC	CACACATGAGCCGAGGAGCAGGTTCTGTGGAGGCGAGAACCC	315	
Db	663	CAGAATATTGAAGATGA	TATGAACAAGTTCAGTCACTTG---ATGAAACAGTATAC	719	
Qy	316	CAGATATCAAGATGAAG	CAACAAAGACAATTCAGTCCCTTCTCCATGAATGGTACAC	375	
Db	720	TCAGAACCTGGAGAGA-	ACCTGCC--CAAGAACCCAGAGGACCAGCAGGAAGAGCTGCAG	776	
Qy	376	GCAGAACATGTTGAGG	GAGAAGACTTGCACAAAGAACAGATGGACCCACAGGAGAACACAA	435	
Db	777	CCGACAGACCATGCTTT	TGT-A-GGA-AGTGATGCAGATCAGACATACGACACCATGGGA	833	
Qy	436	CAAGAGGATGATGAGTT	CTTATGGCCACTGATGTAGATGATGATTTGAGACCCTGGAA	495	
Db	834	ACTGGAGCCGTTGATGA	AGAACTGAAGATAGCTACCATATAGAAGACAGCAGTTCGCCA	893	
Qy	496	CTTGAATATCTCATGA	AGAAACCGGACATAGTTTACCCTCGTGAAGAGACAGTTTCACAA	555	
Db	894	GCCTACAGTCAGATATG	GAGACATGATGTATGAGCAGGAGAAATCCAGATTCCAGGTGAA	953	
Qy	556	GACTGTAAATCAGGAT	TGGAAGAGATGATCTCTGAGCAGGAGAAATCCAGATTCCAGGTGAA	615	
Db	954	CCAGTAGTAGATGA	970		
Qy	616	CCAGTAGTAGAAGATGA	632		







TITLE Structure of the cysteine proteinase (Cp1) gene of *Drosophila melanogaster* and associated mutational effects

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 10772)

AUTHORS Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.

TITLE Direct Submission

JOURNAL Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia

FEATURES

Location/Qualifiers

1..10772

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

join(872..1000,2310..2426,6476..6690,6751..7707)

/gene="Cp1"

/product="cysteine protease"

872..7707

/gene="Cp1"

872..1000

/gene="Cp1"

/number=1

1001..2309

/gene="Cp1"

2310..2426

/gene="Cp1"

/number=2

join(2328..2426,6476..6690,6751..7462)

/gene="Cp1"

/codon\_start=1

/product="cysteine proteinase-1"

/protein\_id="AAB65749.1"

/db\_xref="PID:g2305221"

/db\_xref="GI:2305221"

/translation="MRTAVLLPLLALLAQAQVAFADVVMEEWHTFKLEHRKNYQDET  
EERFLKIPFNENKHKIAKNQRFAGKVSFKLVNRYADLLHFEFLQMGFNFTLHK  
QLRAADESKGVTFFISPAHVTLPKSVDRTKGAVTAVKQGHGCSGWFSSGTALGEG  
HFRKSGVLVSLSPNLDVCTKYGNNGCNGGLMDNAPRKIKONGGIDTDEKSYPEALD  
DSCFNKGTGDRGTDPQDEKMAEAVATGVPVSAIDASHESQFTSEGVYN  
EQCDQKQNDHGVGLVGVFGDESGEDYILVKNWSGVTGDKGFIKRLNRNKGCGIAS  
ASSYPLV"

2427..6475

/gene="Cp1"

4546..4553

/gene="Cp1"

/note="insertion site of P[Casper]{50C}"

/citation=[1]

6476..6690

/gene="Cp1"

/number=3

6691..6750

/gene="Cp1"

6751..7707

/gene="Cp1"

/number=4

join(<8110..9300,9370..>9532)

/product="phenylalanyl tRNA synthetase"

join(8110..9300,9370..>9532)

/note="potential orf"

/codon\_start=1

/product="phenylalanyl tRNA synthetase"

/protein\_id="AAB65750.1"

/db\_xref="PID:g2305222"

/db\_xref="GI:2305222"

/translation="MLLTIRVQGARHLKSTRCLASSAAPKSPSPQLEVSGSTYA  
TGQWNTVTKILSYGVGANKHSLDHPLSIIIRIVNYFYGAYRNORGNPFLFSYVDOMN  
PVVTVOONEDNLLIPADHVSROKSDCYIINQOHLRAHTAHQVELISGLDNFLVVG  
EYVRDEIDSTHPVPHQADAVRLVTKDLFERNPGLELFEETWSGITLADPKLILPH  
SWTKPNPATRAVKLMHEMKHLVLGTLKDLFGPRIKIRWDTYFPFTQPSWELEI  
YFKDNLFVGGCIMRHEILQSRGVHQSIGYAFGVGLERLAWLFDIPRILFVNSDS  
YFKDNLFVGGCIMRHEILQSRGVHQSIGYAFGVGLERLAWLFDIPRILFVNSDS  
DMVEQISLVKDKFKPKTKGSSVCFRIVYRHMERTLTQAEVNEIHKQIASVDSFNVQ  
IX"

2929 a 2357 c 2282 g 3046 t 158 others

BASE COUNT

ORIGIN

Query Match 2.0%; Score 48; DB 21; Length 10772;

Best Local Similarity 21.2%; Pred.No. 3.99e-10;

Matches 32; Conservative 80; Mismatches 34; Indels 5; Gaps 4;

Db 1685 YTKTAAARWYIAWTITMMKMMWTTWKWAMKTYR-TWNMKMYWTSRTTTSAMWY 1743

QY 1365 CTGACACTGATATCTTATCAAAATTTCTACATTAGCTTTAAGTGTTCAGATTACACTT 1424

Db 1744 TWSTWTYKMYAYMKMWRTWARMASWARKWKTSAASAAAYSAWKMWMWAYRAMK 1803

QY 1425 TTGAACACTT-TGTAGCTTT--TAGCTGATTAATAGAAAATAATATTTTCAGTGAAG 1481

Db 1804 KTNWAAWKWR-WKAANWTRWYNTTAA 1833

QY 1482 TTTTAAATTCATTTATTTATTTTAA 1512

RESULT 10

LOCUS AF012089 10772 bp DNA INV 05-AUG-1997

DEFINITION *Drosophila melanogaster* cysteine proteinase-1 (Cp1) gene, complete cds, and phenylalanyl tRNA synthetase gene, partial cds.

ACCESSION AF012089

NID G2305220

VERSION AF012089.1 GI:2305220

KEYWORDS fruit fly.

SOURCE *Drosophila melanogaster*

ORGANISM *Drosophila melanogaster*

REFERENCE 1 (bases 4546 to 4553)

AUTHORS Gray, Y.H., Tanaka, M.M. and Sved, J.A.

TITLE P-element-induced recombination in *Drosophila melanogaster*: hybrid element insertion

JOURNAL Genetics 144 (4), 1601-1610 (1996)

MEDLINE 97132596

REFERENCE 2 (bases 1 to 10772)

AUTHORS Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.

TITLE Structure of the cysteine proteinase (Cp1) gene of *Drosophila melanogaster* and associated mutational effects

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 10772)

AUTHORS Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.

TITLE Direct Submission

JOURNAL Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia

FEATURES

Location/Qualifiers

1..10772

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

join(872..1000,2310..2426,6476..6690,6751..7707)

/gene="Cp1"

/product="cysteine protease"

872..7707

/gene="Cp1"

872..1000

/gene="Cp1"

/number=1

1001..2309

/gene="Cp1"

2310..2426

/gene="Cp1"

/number=2

join(2328..2426,6476..6690,6751..7462)

/gene="Cp1"

/codon\_start=1

/product="cysteine proteinase-1"

/protein\_id="AAB65749.1"

/db\_xref="PID:g2305221"

/db\_xref="GI:2305221"

/translation="MRTAVLLPLLALLAQAQVAFADVVMEEWHTFKLEHRKNYQDET  
EERFLKIPFNENKHKIAKNQRFAGKVSFKLVNRYADLLHFEFLQMGFNFTLHK

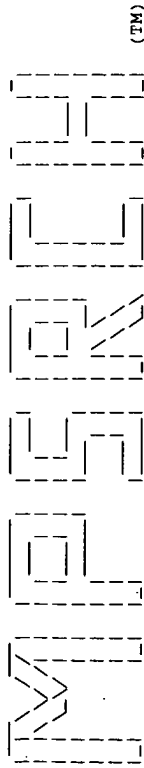




Query Match 1.3%; Score 32; DB 25; Length 215;  
Best Local Similarity 17.9%; Pred. No. 8.35e-02;  
Matches 29; Conservative 66; Mismatches 64; Indels 3; Gaps 3;  
Db 4 VTMSSSVVSRATASCNDAKAKDGNITSSWTTDCNRTWGVCDTDTYRVNNDSGHNKYSS 63  
Cp 1832 ATAGCAGCAGCATCGGCTCTGTGAGCTTTTC-ACACA-TCCTCAATCAACCCAGCTC 1775  
Db 64 ANYNYGNNVGAATHYYHTNVSGADSKTVTDSYNASGTSSSNGTGDGNRSGADSYGSS 123  
Cp 1774 AAGCGTTACTGAGTTCATCTAGTAGAGCACTGAGCGGTCTCTAGTGTCTG-AGGGAA 1716  
Db 124 KTAMTSRNRGTGTANNAYDSRNMGDASYGSDKNTKKHAKNSA 165  
Cp 1715 TTACTCAATTAGCCAAAGGCAGACAATCTTTAAGTTATGCA 1674

Search completed: Fri Oct 22 01:10:25 1999  
Job time : 7554 secs.

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Fri Oct 22 03:09:05 1999; MasPar time 526.27 Seconds  
Tabular output not generated. 994.263 Million cell updates/sec

Title: >US-09-040-485-1  
Description: (1-2442) from US09040485.seq  
Perfect Score: 2442  
N.A. Sequence: 1 CGGGAGCTTGAGGACACAA.....GGTTAACTTAAATATTTT 2442  
Comp: GCCCTCGAAGCTTCCTGCTT.....CCAATTGAAATTATATAAA

Scoring table:  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40 41:part41 42:part42 43:part43  
44:part44 45:part45 46:part46 47:part47 48:part48  
49:part49 50:part50 51:part51 52:part52 53:part53  
54:part54 55:part55 56:part56 57:part57 58:part58  
59:part59 60:part60

Statistics: Mean 9.876; Variance 6.867; scale 1.438

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	289	11.8	316	20	Human gene signature	6.30e-151
2	84	3.4	86	24	Probe 03F1 isolated f	6.12e-30
3	44	1.8	204	1	Base substituted E.co	1.29e-08
4	42	1.7	204	1	Base substituted E.co	1.29e-08
5	39	1.6	91	9	Oligonucleotide probe	3.60e-06
6	39	1.6	91	9	Oligonucleotide probe	3.60e-06
7	38	1.6	114	12	Oligonucleotide probe	1.09e-05
8	37	1.5	114	12	Generic DNA sequence	3.24e-05
9	35	1.4	91	46	Mammalian DNA replica	2.81e-04

C	10	34	1.4	114	12	Q70468	Generic DNA sequence	8.16e-04
C	11	34	1.4	114	12	Q70469	Generic DNA sequence	8.16e-04
C	12	34	1.4	114	12	Q70466	Generic DNA sequence	8.16e-04
C	13	31	1.3	91	46	V44650	Mammalian DNA replica	1.86e-02
C	14	31	1.3	114	12	Q70467	Generic DNA sequence	1.86e-02
C	15	31	1.3	114	12	Q70470	Generic DNA sequence	1.86e-02
C	16	30	1.2	114	12	Q70469	Generic DNA sequence	5.16e-02
C	17	30	1.2	114	12	Q70472	Generic DNA sequence	5.16e-02
C	18	30	1.2	114	12	Q70470	Generic DNA sequence	5.16e-02
C	19	29	1.2	114	12	Q70465	Generic DNA sequence	1.41e-01
C	20	29	1.2	114	12	Q70468	Generic DNA sequence	1.41e-01
C	21	30	1.2	114	12	Q70472	Generic DNA sequence	5.16e-02
C	22	29	1.2	168	32	T76270	Human MDNCF antisense	1.41e-01
C	23	29	1.2	172	32	T76363	Human interleukin 8 a	1.41e-01
C	24	29	1.2	178	32	T76405	Human endothelin-1 an	1.41e-01
C	25	29	1.2	498	3	N50034	Sequence encoding new	1.41e-01
C	26	30	1.2	5746	24	T09225	Partial sequence of v	5.16e-02
C	27	28	1.1	68	34	T73397	Oligonucleotide tag c	3.79e-01
C	28	28	1.1	68	32	T63255	Messenger RNA primer	3.79e-01
C	29	28	1.1	70	24	T14325	Conjugate formed by 1	3.79e-01
C	30	28	1.1	78	59	V82092	D8Y746 yeast cell pol	3.79e-01
C	31	28	1.1	114	12	Q70466	Generic DNA sequence	3.79e-01
C	32	28	1.1	114	12	Q70473	Generic DNA sequence	3.79e-01
C	33	27	1.1	501	3	N50026	Sequence encoding new	1.00e+00
C	34	27	1.1	501	3	N50023	Sequence encoding new	1.00e+00
C	35	27	1.1	501	3	N50026	Sequence encoding new	1.00e+00
C	36	27	1.1	501	3	N50030	Sequence encoding new	1.00e+00
C	37	27	1.1	501	3	N50029	Sequence encoding new	1.00e+00
C	38	27	1.1	501	3	N50033	Sequence encoding new	1.00e+00
C	39	27	1.1	779	60	X00240	Murine osteogenic pro	1.00e+00
C	40	27	1.1	779	40	V15215	Murine osteogenic pro	1.00e+00
C	41	27	1.1	779	12	Q72702	Mature mouse CBMP3 c	1.00e+00
C	42	27	1.1	779	18	T02604	Mature murine CBMP3 c	1.00e+00
C	43	27	1.1	780	9	Q53152	Sequence encoding mur	1.00e+00
C	44	27	1.1	1189	7	Q40203	Sequence of JFSig2 DN	1.00e+00
C	45	27	1.1	2277	41	V05370	Human telomerase p105	1.00e+00

ALIGNMENTS

RESULT 1  
ID T23982 standard; cDNA to mRNA; 316 BP.  
AC T23982:  
DE T2-SEP-1996 (first entry)  
DE Human gene signature HUMGS05943.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 1501; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SQ Sequence 316 BP; 107 A; 35 C; 62 G; 105 T; 1; Gaps 1; Mismatches 308; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

Query Match 11.8%; Score 289; DB 20; Length 316;  
 Best Local Similarity 96.9%; Pred. No. 6.30e-151;  
 Matches 308; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

Db 1 gatcatagttattatacaatgtagtgagctgcctgcagctgggtctgcagtgatgtaataaacc 60  
 QY 2125 GATCATAGTTATTATACAAATGTAGTGCCTGCATGGGTACTCGATGTCTAATGAAACC 2184

Db 61 tgaataataaa---gataataaagaagaacaataattttctaaagctgtgctgctgtgat 117  
 QY 2185 TGAATAATAAATAAGATAATAAGAAAGCAATAATTTCTFAAGCTGTGCTGCGGTGAT 2244

Db 118 acagagatgatactcaaatataataaaactcttcattttgtgaattatagaagctactn 177  
 QY 2245 ACAGAGATGATACTCAATATATATAAACTCTTCATTTGTGAATATATAGAGCTACTT 2304

Db 178 nmntaaagccatttttttagggaaactaaagagtgacatagactgatgaatgagca 237  
 QY 2305 TTTATAAGCCATATTTTATAGGAACTAAGAGGTGACATAGAACTGATGATGAGTA 2364

Db 238 aaagtgaagtttctgagattttgtgagaactctggacgttgagattcattatgctgtgg 297  
 QY 2365 AAGTGAAGTTTCTGCTGGATTTTGTGAAGCTGGAAGCTTGGAGCTTCATTTATGCTGTG 2424

Db 298 ttaactttaaatatttt 315  
 QY 2425 TTAACTTTAAATATTTT 2442

RESULT 2  
 ID T28082 standard; DNA; 86 BP.  
 AC T28082.  
 DT 31-DEC-1996. (first entry)  
 DE Probe 03F1 isolated from fibroblasts.  
 KW Polymerase chain reaction; PCR; primer; amplification; fibroblast; AIDS;  
 KW enhanced differential display; EDD; mRNA preparation; senescent cell;  
 KW quiescent cell; dividing cell; senescence-related gene; gene expression;  
 KW non-senescent cell; age-related lipofuscin; retina; therapy; liver spot;  
 KW donor tissue; senescent melanocyte; melanin; hypopigmentation; ss.  
 PS Synthetic.  
 PN WO9613610-A2.  
 PD 09-MAY-1996.  
 PF 24-AUG-1995; U11230.  
 PR 31-OCT-1994; US-332420.  
 PA (GERO-) GERON CORP.  
 PI Feng J, Funk W, Hirsch KS, Linskens MHK, Villeponteau B;  
 PI West MD;  
 DR WPI: 96-251464/25.  
 PT Identifying, isolating and regulating senescence-related genes -  
 PT useful to ameliorate problems associated with accumulation of  
 PT senescent cells, e.g. age-related lipofuscin accumulation in the  
 PT retina and AIDS  
 PS Claim 8; Page 36; 135pp; English.  
 CC T28076-T28113, and T28131-T28173 represent novel senescent-related gene  
 CC sequences isolated from fibroblasts using the method of the invention.  
 CC In the method of the invention, mRNA is isolated from a senescent cell,  
 CC and a young quiescent cell, and the mRNAs are amplified (using primers  
 CC such as those shown in T28044-T28075) in separate reaction mixtures. The  
 CC amplified sequences are then separated by size or charge, and the  
 CC products are analysed to identify a gene from young quiescent cells and  
 CC dividing cells, that is present at a different level from senescent  
 CC cells. To enhance the method even more, it can be performed in  
 CC conjunction with an enhanced differential display (EDD) method (an mRNA  
 CC preparation method) on the fibroblasts. The method can be used for the  
 CC rapid and efficient identification and isolation of senescence-related  
 CC genes and gene products, and to detect and distinguish between senescent  
 CC and non-senescent cells. It can also be used to destroy cells expressing

CC senescence specific (or related) gene products, and to screen for  
 CC compounds capable of altering gene expression in senescent cells. The  
 CC method can also be used to ameliorate problems associated with the  
 CC accumulation of senescent cells such as age-related lipofuscin  
 CC accumulation in the retina, and in the treatment of AIDS. Also, the  
 CC method can be used to distinguish young cells from senescent cells in  
 CC donor tissue, which is useful in removing senescent melanocytes  
 CC overexpressing melanin which cause hypopigmentation, or liver spots.  
 SQ Sequence 86 BP; 33 A; 10 C; 13 G; 30 T;

Query Match 3.4%; Score 84; DB 24; Length 86;  
 Best Local Similarity 98.8%; Pred. No. 6.12e-30;  
 Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 ataataaaactcttcattttgcgaattatagaagctactttttataagccatattttt 60  
 QY 2265 ATAATAAACTCTTCATTTGTGAATATATAGAGCTACTTTTATAAGCCATATTTT 2324

Db 61 tagggaactaaagagtgacatagaa 86  
 QY 2325 TAGGGAACCTAAGAGGTGACATAGAA 2350

RESULT 3  
 ID N81164 standard; DNA; 204 BP.  
 AC N81164.  
 DT 08-NOV-1990 (first entry)  
 DE Base substituted E.coli beta-galactosidase alpha-fragment.  
 KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.  
 OS Escherichia coli.  
 FH Key Location/Qualifiers  
 FT misc\_feature 19..69  
 FT primer\_bind /tag= a  
 FT /function=multiple cloning site  
 FT 187..204  
 FT /tag= b  
 PN BP-285123-A.  
 PD 05-MAY-1988.  
 PF 30-MAR-1988; 105163.  
 PR 03-APR-1987; US-034819.  
 PA (SUSO) SUOMEN SOKERI OY.  
 PI Lehtovaara P, Kowles J, Koivula A, Bamford J, Reinikainen T;  
 DR WPI: 88-279927/40.  
 PT Introducing random point mutations into nucleic acids -  
 PT by prepn of single stranded template, annealing a primer, elongation,  
 PT misincorporation, completion of molecules and screening.  
 PS Disclosure; P; English.  
 CC Random point mutations were introduced into the alpha fragment of  
 CC E.coli beta-galactosidase. The wild type sequence was obtained as a  
 CC single stranded template and an oligonucleotide was hybridised to  
 CC it to generate a popn of DNA molecules which terminate at all  
 CC possible nucleotide positions within a specified region. The  
 CC variable 3' ends generated in this way are used as primers for  
 CC reverse transcriptase. Nucleotides are misincorporated by the  
 CC transcriptase and the molecules are completed to forms that can be  
 CC amplified and then expressed in a suitable host-vector system.  
 CC The sequence covers all 176 diff base substitutions, most of which  
 CC occurred singularly in any given mutant.  
 CC See also P80575.  
 SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 1.8%; Score 44; DB 1; Length 204;  
 Best Local Similarity 17.2%; Pred. No. 1.29e-08;  
 Matches 23; Conservative 59; Mismatches 52; Indels 0; Gaps 0;

Db 66 gacycgaaycchvcgcgymttthrrmrbrnrvdyrnsdaaawccyrrsvkydc 125  
 QY 228 GAGATCTACTTACAGCCAGCTCCGCCAGAGAGGCTGAGCCACACACTGAGCCCGA 287

Db 126 cynacthddhyvbbvynvnhnncnccbnhvhvbnhnrnwayvrhddrddvh 185  
 QY 288 GGAGCAGGTTCTCTGTGGAGGCAGAACCCCGAGATATCGAGATGAGCAAGAACAAAT 347

```

PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 1.6%; Score 39; DB 9; Length 91;
Best Local Similarity 0.8%; Pred. No. 3.60e-06;
Matches 0; Conservative 44; Mismatches 5; Indels 0; Gaps 0;

Db 12 svhsyvvvvhvshhshvhhvshvsvvvvhhvhhvhhvhhvhhvhhvsv 60
QY 813 CGAGGAAGTACCACGAGACTACTAAAGCTTCAAAAGAGCTGCCCTACC 861

RESULT 6
ID Q51745 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
PR WPI: 93-378844/48.
PT New oligo-nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 1.6%; Score 39; DB 9; Length 91;
Best Local Similarity 5.6%; Pred. No. 3.60e-06;
Matches 3; Conservative 46; Mismatches 4; Indels 1; Gaps 1;

Db 12 svhsyvvvvhvshhshvhhvshvsvvvvhhvhhvhhvhhvhhvsvctca 64
Cp 127 CAAGATCAACCAACGACAGCTACAGATGTCGAGAGCGCCACCAATGCAATCA 74

RESULT 7
ID Q70467 standard; DNA; 114 BP.
AC Q70467;
DE 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSNR peptide library.
KW TSNAR; totally synthetic affinity reagent; Synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT FT /tag= a
FT FT /note= "this sequence represents '2'; 2 can be a
FT FT sequence of 6, 9 or 12 nucleotides (see
FT FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.

```











\*\*\*\*\*

WORLD

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPerch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 03:21:25 1999; MasPar time 174.29 Seconds  
Tabular output not generated. 1211.823 Million cell updates/sec

Title: >US-09-040-485-1  
Description: (1-2442) from US09040485.seq  
Perfect Score: 2442  
N.A. Sequence: 1 CGGGAGCTTGAAGGACACAA.....GGTTAACTTTAAATATTTT 2442  
Comp: GCCTCGAAGCTTCTGTGTT.....CCAATTGAAATTTATAAANA

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-issued  
1:5A\_COMB 2:5B\_COMB 3:5C\_COMB 4:PCF9\_COMB 5:backfiles1

Statistics: Mean 9.345; Variance 5.512; scale 1.695

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
c 1	108	4.4	7218	2	US-08-232- Sequence 14, Applicati	4.94e-53
c 2	84	3.4	86	2	US-08-332- Sequence 39, Applicati	5.69e-37
c 3	51	2.1	7218	2	US-08-232- Sequence 14, Applicati	7.39e-16
c 4	42	1.7	215	1	US-08-238- Sequence 5, Applicatio	1.83e-10
c 5	42	1.7	965	3	US-08-388- Sequence 22, Applicati	1.83e-10
c 6	32	1.3	215	1	US-08-238- Sequence 5, Applicatio	8.08e-05
c 7	30	1.2	965	3	US-08-388- Sequence 22, Applicati	9.42e-04
c 8	26	1.1	54	3	US-08-452- Sequence 15, Applicati	1.05e-01
c 9	26	1.1	54	3	US-08-452- Sequence 14, Applicati	1.05e-01
c 10	27	1.1	90	3	US-08-442- Sequence 30, Applicati	3.33e-02
c 11	27	1.1	779	3	US-07-841- Sequence 22, Applicati	3.33e-02
c 12	27	1.1	779	3	US-08-449- Sequence 22, Applicati	3.33e-02
c 13	27	1.1	779	3	US-08-447- Sequence 22, Applicati	3.33e-02
c 14	27	1.1	779	1	US-08-147- Sequence 22, Applicati	3.33e-02
c 15	27	1.1	1189	1	US-07-781- Sequence 4, Applicatio	3.33e-02
c 16	27	1.1	1189	4	PCT-US92-0 Sequence 23, Applicati	1.05e-01
c 17	26	1.1	2117	2	US-08-431- Sequence 2, Applicatio	3.33e-02
c 18	27	1.1	2277	2	US-08-676- Sequence 2, Applicatio	3.33e-02
c 19	27	1.1	2277	3	US-08-676- Sequence 2, Applicatio	3.33e-02
c 20	24	1.0	42	3	US-08-452- Sequence 43, Applicati	9.92e-01

c 21	24	1.0	74	4	PCT-US95-1	Sequence 94, Applicati	9.92e-01
c 22	24	1.0	74	4	PCT-US95-1	Sequence 100, Applicat	9.92e-01
c 23	24	1.0	75	4	PCT-US95-1	Sequence 99, Applicati	9.92e-01
c 24	24	1.0	81	4	PCT-US95-1	Sequence 92, Applicati	9.92e-01
c 25	24	1.0	81	4	PCT-US95-1	Sequence 98, Applicati	9.92e-01
c 26	24	1.0	82	4	PCT-US95-1	Sequence 97, Applicati	9.92e-01
c 27	24	1.0	242	1	US-08-273- Sequence 1, Applicatio	9.92e-01	9.92e-01
c 28	25	1.0	377	3	US-08-332- Sequence 1, Applicatio	9.92e-01	9.92e-01
c 29	24	1.0	630	1	US-08-375- Sequence 10, Applicati	9.92e-01	9.92e-01
c 30	24	1.0	630	4	PCT-US92-0	Sequence 10, Applicati	9.92e-01
c 31	24	1.0	1020	4	PCT-US95-0	Sequence 10, Applicati	9.92e-01
c 32	24	1.0	2504	1	US-08-484- Sequence 15, Applicati	9.92e-01	9.92e-01
c 33	24	1.0	2504	1	US-08-484- Sequence 15, Applicati	9.92e-01	9.92e-01
c 34	24	1.0	2504	1	US-08-484- Sequence 15, Applicati	9.92e-01	9.92e-01
c 35	24	1.0	2504	1	US-08-484- Sequence 15, Applicati	9.92e-01	9.92e-01
c 36	24	1.0	2666	1	US-08-369- Sequence 3, Applicatio	9.92e-01	9.92e-01
c 37	24	1.0	6877	1	US-08-347- Sequence 1, Applicatio	9.92e-01	9.92e-01
c 38	24	1.0	8082	4	PCT-US93-0	Sequence 28, Applicati	9.92e-01
c 39	24	1.0	8082	2	US-08-187- Sequence 1, Applicatio	9.92e-01	9.92e-01
c 40	24	1.0	8082	2	US-08-306- Sequence 41, Applicati	9.92e-01	9.92e-01
c 41	25	1.0	9636	2	US-08-323- Sequence 1, Applicatio	3.27e-01	3.27e-01
c 42	23	0.9	75	4	PCT-US95-1	Sequence 99, Applicati	2.94e+00
c 43	23	0.9	81	4	PCT-US95-1	Sequence 92, Applicati	2.94e+00
c 44	23	0.9	82	4	PCT-US95-1	Sequence 97, Applicati	2.94e+00
c 45	23	0.9	2377	3	US-08-967- Sequence 26, Applicati	2.94e+00	2.94e+00

ALIGNMENTS

RESULT 1  
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.  
AC xxxxxx  
DT  
DE Sequence 14, Application US/08232463  
CC Sequence 14, Application US/08232463  
CC Patent No. 5670367  
CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.  
CC APPLICANT: SCHEIFLINGER, F.  
CC APPLICANT: FALKNER, F. G.  
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
CC NUMBER OF SEQUENCES: 52  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 1800 Diagonal Road, Suite 500  
CC CITY: Alexandria  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22313-0299  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA: US/08/232.463  
CC APPLICATION NUMBER: US/07/935.313  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: EP 91 114 300.6  
CC FILING DATE: 26-AUG-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: BENT, Stephen A.  
CC REGISTRATION NUMBER: 29,768  
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703)836-9300  
CC TELEFAX: (703)683-4109  
CC TELEX: 899149  
CC INFORMATION FOR SEQ ID NO: 14:  
CC SEQUENCE CHARACTERISTICS:



CC LENGTH: 7218 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC IMMEDIATE SOURCE:  
CC CLONE: pTZgpt-fls  
SQ SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

Query Match 2.1%; Score 51; DB 2; Length 7218;  
Best Local Similarity 1.1%; Pred. No. 7.39e-16;  
Matches 2; Conservative 112; Mismatches 63; Indels 0; Gaps 0;

Db 1065 ATYY 1124  
QY 979 ATTATAGTCCACGCGCATTTTAGTGATTTTCTTTTCTTTTGAACACACATTTCTAA 1038  
Db 1125 YYY 1184  
QY 1039 AATGTGATGTTACATCTCGCATGTCCTTTTGATAGCATTAGTGATCCATTGGATTT 1098  
Db 1185 YYY 1241  
QY 1099 CTTTTTCTTTTGAGACAGCTTTTAGTCTTACCTGAATTATGTTGTTTTC 1155

RESULT 4  
ID US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.  
AC xxxxxx  
DT  
DE Sequence 5, Application US/08238163  
CC Sequence 5, Application US/08238163  
CC Patent No. 5569830  
CC GENERAL INFORMATION:  
CC APPLICANT: BENNETT, Alan  
CC APPLICANT: LABAVITCH, John M.  
CC APPLICANT: POWELL, Ann  
CC APPLICANT: STOTZ, Henrik  
CC TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL  
CC NUMBER OF SEQUENCES: 24  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Townsend and Townsend Kourie and Crew  
CC STREET: Steuart Street Tower, One Market Plaza  
CC CITY: San Francisco  
CC STATE: California  
CC COUNTRY: US  
CC ZIP: 94105-1493  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/238,163  
CC FILING DATE: 03-MAY-1994  
CC CLASSIFICATION: 800  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Bastian, Kevin L.  
CC REGISTRATION NUMBER: 34,774  
CC REFERENCE/DOCKET NUMBER: 2307E-540  
CC TELEPHONE: (415) 543-9600  
CC TELEFAX: (415) 543-5043  
CC INFORMATION FOR SEQ ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 215 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC FEATURE:  
CC NAME/KEY: misc\_feature  
CC LOCATION: 1..215

US-09-040-485-1.l.rni

CC OTHER INFORMATION: /standard\_name= "Deduced amino acid  
CC OTHER INFORMATION: sequence of PGIP from bean."  
SQ SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

Query Match 1.7%; Score 42; DB 1; Length 215;  
Best Local Similarity 18.6%; Pred. No. 1.83e-10;  
Matches 37; Conservative 77; Mismatches 82; Indels 3; Gaps 3;

Db 6 MSSSVVSRATSCNDKAKDGNNTSSWTTCCNRTGWCDTDTTYRVNNDGHNKYSAN 65  
QY 430 CCACCAAGAGAGATGATGAGTTTCTTATGCGGACTAGATAGATAGATTGAGACC 489

Db 66 YNKGNNV-GAAATHYYTH-TNVSADSKVTDSYNASGTSSSNGGTDGNSRGADSYSS 123  
QY 490 CTGGAACCTGAATGATCTCATGAAGAACCCAGCATAGTACCACGTGGAAGACACATT 549

Db 124 KTAMTSNRRTGKTANNVDS-RNMGDASVGS DKNTKKHAKNSADGKVGSKNGNDRNRYG 182  
QY 550 TCACAAGACTGTAATCAGGATATGGAAGAGATGATGCTGTGAGCAGGAAAAATCCAGATTCC 609

Db 183 TGTKSNNVNNCGGKNKRDV 201  
QY 610 AGTGAACCACTAGTAGAAG 628

RESULT 5  
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.  
AC xxxxxx  
DT  
DE Sequence 22, Application US/08388672A  
CC Sequence 22, Application US/08388672A  
CC Patent No. 5795961  
CC GENERAL INFORMATION:  
CC APPLICANT: Wallace, T. Paul  
CC APPLICANT: Harris, William J.  
CC APPLICANT: Carr, Frank J.  
CC APPLICANT: Old, Lloyd J.  
CC APPLICANT: Welt, Sydney  
CC APPLICANT: Kitamura, Kunio  
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B  
CC TITLE OF INVENTION: Antibodies  
CC NUMBER OF SEQUENCES: 25  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Felfe and Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: U.S.A.  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/388,672A  
CC FILING DATE: 14-FEB-1995  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hanson, No. 5795961man D.  
CC REGISTRATION NUMBER: 30,946  
CC REFERENCE/DOCKET NUMBER: LUD 5409  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 212-688-9200  
CC TELEFAX: 212-838-3884  
CC INFORMATION FOR SEQ ID NO: 22:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 965 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

Query Match 1.7%; Score 42; DB 3; Length 965;  
Best Local Similarity 23.9%; Pred. No. 1.83e-10;  
Matches 34; Conservative 59; Mismatches 48; Indels 1; Gaps 1;

Db 759 AGCTGGAATCAAAATGGGTTKURHUVHVGVRSTSTCTASDYTTSYWVWVRGWD 818

Qy 5 AGCTTGAAGGACAGAAGATGGAGAAAGCGGACTCTCAGGAACCTTCATCTTC-ACG 63

Db 819 YGGVTYNGRGRVTMADTSSNSRSTVADTANYVCVRGRSYSDGDYGGTGVTVS 878

Qy 64 TGGTTATGGTGAATGCATGCTGGCGGCTGGACATCTGTAGCTGCTGCTGGTTGAT 123

Db 879 SHUKDMTSSSSASVGRDVTTC 900

Qy 124 CTGTGTGACTATGAGGAAGTTC 145

RESULT 6  
ID US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.  
AC xxxxxx

DE Sequence 5, Application US/08238163

CC Sequence 5, Application US/08238163

CC Patent No. 5569830

CC GENERAL INFORMATION:

CC APPLICANT: BENNETT, Alan

CC APPLICANT: LABAVITCH, John M.

CC APPLICANT: FOWELL, Ann

CC APPLICANT: STOTZ, Henrik

CC TITLE OF INVENTION: POLY INHIBITORS OF FUNGAL

CC TITLE OF INVENTION: POLY GALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE

CC NUMBER OF SEQUENCES: 24

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Townsend and Townsend Kourie and Crew

CC STREET: Steuart Street Tower, One Market Plaza

CC CITY: San Francisco

CC STATE: California

CC COUNTRY: US

CC ZIP: 94103-1493

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent in Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/238,163

CC FILING DATE: 03-MAY-1994

CC CLASSIFICATION: 800

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Bastian, Kevin L.

CC REGISTRATION NUMBER: 34,774

CC REFERENCE/DOCKET NUMBER: 2307E-540

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (415) 543-9600

CC TELEFAX: (415) 543-5043

CC INFORMATION FOR SEQ ID NO: 5:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 215 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: single

CC TOPOLOGY: unknown

CC MOLECULE TYPE: protein

CC FEATURE:

CC NAME/KEY: misc.feature

CC LOCATION: 1..215

CC OTHER INFORMATION: /standard\_name= "Deduced amino acid

CC OTHER INFORMATION: sequence of PGIP from bean."

CC SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

Qy SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

Query Match 1.3%; Score 32; DB 1; Length 215;

Best Local Similarity 17.9%; Pred. No. 8.08e-05;

Matches 29; Conservative 56; Mismatches 64; Indels 3; Gaps 3;

Db 4 VTMSSSVVSRRTASNDKAKDGNNTSSWTTDCNRTGWVCDTDTTYRVNDSGHNKYSS 63

Qp 1832 ATAGCAGCAGCATCGGGCTCTGTGAGCTTTTC-ACACA-TCCATCAATCAACCCAGCTC 1775

Db 64 ANYNNGNNVGAATHTYHTNVSAGDSKTVTDSYNASGTSNNGTGDGNSRGSADSYGSS 123

Qp 1774 AAGCGGTTACTGAGTTCATCTAGTAGACACTGACGGCTCTCTAGTGCTGG-AGGGGAA 1716

Db 124 KTAMTSRNTGTANNVDSRNMGDSVGSDDKNTKKHAKNSA 165

Qp 1715 TTACTCAATTAGCAAGGAGCAATAATCTTTAAGTTATGCA 1674

RESULT 7  
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.  
AC xxxxxx

DE Sequence 22, Application US/08388672A

CC Sequence 22, Application US/08388672A

CC Patent No. 5795961

CC GENERAL INFORMATION:

CC APPLICANT: Wallace, T. Paul

CC APPLICANT: Harris, William J.

CC APPLICANT: Carr, Frank J.

CC APPLICANT: Old, Lloyd J.

CC APPLICANT: Weit, Sydney

CC APPLICANT: Kitamura, Kunio

CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B

CC TITLE OF INVENTION: Antibodies

CC NUMBER OF SEQUENCES: 25

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Felfe and Lynch

CC STREET: 805 Third Avenue

CC CITY: New York

CC STATE: New York

CC COUNTRY: U.S.A.

CC ZIP: 10022

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC OPERATING SYSTEM: IBM PC compatible

CC SOFTWARE: Patent in Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/388,672A

CC FILING DATE: 14-FEB-1995

CC CLASSIFICATION:

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Hanson, No. 5795961man D.

CC REGISTRATION NUMBER: 30,946

CC REFERENCE/DOCKET NUMBER: LUD 5409

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 212-688-9200

CC TELEFAX: 212-688-3884

CC INFORMATION FOR SEQ ID NO: 22:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 965 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: unknown

CC TOPOLOGY: unknown

CC MOLECULE TYPE: DNA (genomic)

Qy SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

Query Match 1.2%; Score 30; DB 3; Length 965;

Best Local Similarity 15.9%; Pred. No. 9.42e-04;

Matches 23; Conservative 64; Mismatches 57; Indels 1; Gaps 1;

Db 789 SGGVRSTSTCTASDYTTSY-WGVWVRGWDYGGGYTYNKRGRVTMADTSSNSRST 847

Qp 1958 GGTAGCTACCTCTCTTCTGACAAAGGATATTTTCAACAGAGCAATATTCTTGAGCCC 1899

Db 848 AADTAVYCVGRGVSDDGGDYWGTTTTSVSHUVKDMTSSSSASVGDRTTCRSSTTHG 907





CC COUNTRY: USA  
CC ZIP: 94111-3834  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent in Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/442,461D  
CC FILING DATE: 17-MAY-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Liebeschuetz, Joe  
CC REGISTRATION NUMBER: 37,505  
CC REFERENCE/DOCKET NUMBER: 023070U-060500US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 576-0200  
CC TELEFAX: (415) 576-0300  
CC INFORMATION FOR SEQ ID NO: 30:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 90 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA  
SQ SEQUENCE 90 BP; 12 A; 16 C; 14 G; 20 T; 28 OTHER.

Query Match 1.1%; Score 27; DB 3; Length 90;  
Best Local Similarity 19.5%; Pred. No. 3.33e-02;  
Matches 8; Conservative 26; Mismatches 7; Indels 0; Gaps 0;

Db 19 MRTMR 59  
Cp 202 CATCATCCATCAATCAAAATCTCCATCCATCAGCATCATAG 162

RESULT 11  
ID US-07-841-646-22 STANDARD; DNA; UNC; 779 BP.  
AC xxxxxx

Sequence 22, Application US/07841646  
Sequence 22, Application US/07841646  
Patent No. 5266683  
GENERAL INFORMATION:  
CC APPLICANT: OPPERMAN, HERMANN  
CC APPLICANT: OZKAYNAK, ENGIN  
CC APPLICANT: KUBERASAMPATH, THANGAVEL  
CC APPLICANT: RUEGER, DAVID C.  
CC APPLICANT: PANG, ROY H.L.  
CC TITLE OF INVENTION: OSTEOGENIC DEVICES  
CC NUMBER OF SEQUENCES: 33  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
CC STREET: 53 STATE STREET  
CC CITY: BOSTON  
CC STATE: MASSACHUSETTS  
CC COUNTRY: U.S.A.  
CC ZIP: 02109  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent in Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/841,646  
CC FILING DATE: 19920221  
CC CLASSIFICATION: 530  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 810,560  
CC FILING DATE: 20-DEC-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 827,052  
CC FILING DATE: 28-JAN-1992

CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 660,162  
CC FILING DATE: 22-FEB-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,988  
CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,849  
CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 616,374  
CC FILING DATE: 21-NOV-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 600,024  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 599,543  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 579,865  
CC FILING DATE: 07-SEP-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 569,920  
CC FILING DATE: 20-AUG-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 483,913  
CC FILING DATE: 22-FEB-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 422,613  
CC FILING DATE: 17-OCT-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 315,342  
CC FILING DATE: 23-FEB-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 232,630  
CC FILING DATE: 15-AUG-1988  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 179,460  
CC FILING DATE: 08-APR-1988  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: PITCHER, EDMUND R.  
CC REGISTRATION NUMBER: 27,829  
CC REFERENCE/DOCKET NUMBER: CRP-001CP6  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 617/248-7000  
CC TELEFAX: 617/248-7100  
CC INFORMATION FOR SEQ ID NO: 22:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 779 base pairs  
CC TYPE: NUCLEIC ACID  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA  
CC ORIGINAL SOURCE:  
CC ORGANISM: MURINE  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1..549  
CC OTHER INFORMATION: /product= "MATURE mbmp3"  
SQ SEQUENCE 779 BP; 241 A; 164 C; 188 G; 186 T; 0 OTHER.

Query Match 1.1%; Score 27; DB 1; Length 779;  
Best Local Similarity 76.5%; Pred. No. 3.33e-02;  
Matches 39; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 726 TGAGATTTTAGTAATGTGGATCTCTAAATAAAAAAAAAAAAAAAAAAAAAA 776  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Cp 1046 TGACATTTTGAATGTGTGTCTTCTAAAGAGAGAGAGAGAGAGAGAGAG 996

RESULT 12  
ID US-08-449-700-22 STANDARD; DNA; UNC; 779 BP.  
AC xxxxxx

DT Sequence 22, Application US/08449700  
DE Sequence 22, Application US/08449700  
CC Patent No. 5863758  
CC GENERAL INFORMATION:  
CC APPLICANT: OPPERMAN, HERMANN  
CC APPLICANT: OZKAYNAK, ENGIN  
CC APPLICANT: KUBERASAMPATH, THANGAVEL  
CC APPLICANT: RUEGER, DAVID C.  
CC APPLICANT: PANG, ROY H.L.  
CC TITLE OF INVENTION: OSTEOGENIC DEVICES  
CC NUMBER OF SEQUENCES: 33  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
CC STREET: 53 STATE STREET  
CC CITY: BOSTON  
CC STATE: MASSACHUSETTS  
CC COUNTRY: U.S.A.  
CC ZIP: 02109  
CC COMPUTER READABLE FORM: disk  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/449,700  
CC FILING DATE: 21-FEB-1992  
CC CLASSIFICATION: 530  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 810,560  
CC FILING DATE: 20-DEC-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 827,052  
CC FILING DATE: 28-JAN-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 660,162  
CC FILING DATE: 22-FEB-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,988  
CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,849  
CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 616,374  
CC FILING DATE: 21-NOV-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 600,024  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 599,543  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 579,865  
CC FILING DATE: 07-SEP-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 569,920  
CC FILING DATE: 20-AUG-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 483,913  
CC FILING DATE: 22-FEB-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 422,613  
CC FILING DATE: 17-OCT-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 315,342  
CC FILING DATE: 23-FEB-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 232,630  
CC FILING DATE: 15-AUG-1988  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 179,460  
CC FILING DATE: 08-APR-1988

CC ATTORNEY/AGENT INFORMATION:  
CC NAME: PITCHER, EDMUND R.  
CC REGISTRATION NUMBER: 27,829  
CC REFERENCE/DOCKET NUMBER: CRP-001CP6  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 617/248-7000  
CC TELEFAX: 617/248-7100  
CC INFORMATION FOR SEQ ID NO: 22:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 779 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA  
CC ORIGINAL SOURCE:  
CC ORGANISM: MURINE  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1..549  
CC OTHER INFORMATION: /product= "MATURE mbmp3"  
CC SEQ SEQUENCE 779 BP; 241 A; 184 C; 186 G; 186 T; 0 OTHER.  
  
Query Match 1.1%; Score 27; DB 3; Length 779;  
Best Local Similarity 76.5%; Pred. No. 3.33e-02;  
Matches 39; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
  
Db 726 TGAGATTTTAGTAATGTGATCTCTAAAAAAGAAAAAATACATAAA 776  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Cp 1046 TGACATTTTAGTAATGTGATCTCTAAAAAAGAAAAAATACATAAA 996  
  
RESULT 13  
ID US-08-447-570-22 STANDARD; DNA; UNC; 779 BP.  
AC xxxxxx  
DT  
DE Sequence 22, Application US/08447570  
CC Sequence 22, Application US/08447570  
CC Patent No. 5714589  
CC GENERAL INFORMATION:  
CC APPLICANT: OPPERMAN, HERMANN  
CC APPLICANT: OZKAYNAK, ENGIN  
CC APPLICANT: KUBERASAMPATH, THANGAVEL  
CC APPLICANT: RUEGER, DAVID C.  
CC APPLICANT: PANG, ROY H.L.  
CC TITLE OF INVENTION: OSTEOGENIC DEVICES  
CC NUMBER OF SEQUENCES: 33  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
CC STREET: 53 STATE STREET  
CC CITY: BOSTON  
CC STATE: MASSACHUSETTS  
CC COUNTRY: U.S.A.  
CC ZIP: 02109  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/447,570  
CC FILING DATE: 21-FEB-1992  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 810,560  
CC FILING DATE: 20-DEC-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 827,052  
CC FILING DATE: 28-JAN-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 660,162  
CC FILING DATE: 22-FEB-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,988

CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,849  
CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 616,374  
CC FILING DATE: 21-NOV-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 600,024  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 599,543  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 579,865  
CC FILING DATE: 07-SEP-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 569,920  
CC FILING DATE: 20-AUG-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 483,913  
CC FILING DATE: 22-FEB-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 422,613  
CC FILING DATE: 17-OCT-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 315,342  
CC FILING DATE: 23-FEB-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 232,630  
CC FILING DATE: 15-AUG-1988  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 179,460  
CC FILING DATE: 08-APR-1988  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: PITCHER, EDMUND R.  
CC REGISTRATION NUMBER: 27,829  
CC REFERENCE/DOCKET NUMBER: CRP-001CP6  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 617/248-7000  
CC TELEFAX: 617/248-7100  
CC INFORMATION FOR SEQ ID NO: 22:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 779 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA  
CC ORIGINAL SOURCE:  
CC ORGANISM: MURINE  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1..549  
CC OTHER INFORMATION: /product= "MATURE mbMP3"  
SQ SEQUENCE 779 BP; 241 A; 164 C; 188 G; 186 T; 0 OTHER.

Query Match 1.1%; Score 27; DB 2; Length 779;  
Best Local Similarity 76.5%; Pred. No. 3.33e-02;  
Matches 39; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
Db 726 TCAGATTTTAGTAATGGATCTCTAATAAAAAAAAAAAAAAAAAAAAAA 776  
Cp 1046 TGACATTTTGAAGATGTGTCTTCAAGAGAGAGAGAGAGAGAGAG 996

RESULT 14  
ID US-08-147-023-22 STANDARD; DNB; UNC; 779 BP.  
AC xxxxxx  
DT  
DE Sequence 22, Application US/08147023  
CC Sequence 22, Application US/08147023  
CC Patent No. 5468845  
CC GENERAL INFORMATION:

CC APPLICANT: OPPERMAN, HERMANN  
CC APPLICANT: OZKAYNAK, ENGIN  
CC APPLICANT: KUBERASAMPATH, THANGAVEL  
CC APPLICANT: RUEGER, DAVID C.  
CC APPLICANT: PANG, ROY H.L.  
CC TITLE OF INVENTION: OSTEOGENIC DEVICES  
CC NUMBER OF SEQUENCES: 33  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
CC STREET: 53 STATE STREET  
CC CITY: BOSTON  
CC STATE: MASSACHUSETTS  
CC COUNTRY: U.S.A.  
CC ZIP: 02109  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/147,023  
CC FILING DATE: 21-FEB-1992  
CC CLASSIFICATION: 530  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 810,560  
CC FILING DATE: 20-DEC-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 827,052  
CC FILING DATE: 28-JAN-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 660,162  
CC FILING DATE: 22-FEB-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,988  
CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,849  
CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 616,374  
CC FILING DATE: 21-NOV-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 600,024  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 599,543  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 579,865  
CC FILING DATE: 07-SEP-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 569,920  
CC FILING DATE: 20-AUG-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 483,913  
CC FILING DATE: 22-FEB-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 422,613  
CC FILING DATE: 17-OCT-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 315,342  
CC FILING DATE: 23-FEB-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 232,630  
CC FILING DATE: 15-AUG-1988  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 179,460  
CC FILING DATE: 08-APR-1988  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: PITCHER, EDMUND R.  
CC REGISTRATION NUMBER: 27,829  
CC REFERENCE/DOCKET NUMBER: CRP-001CP6  
CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 617/248-7000  
CC TELEFAX: 617/248-7100  
CC INFORMATION FOR SEQ ID NO: 22:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 779 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA  
CC ORIGINAL SOURCE:  
CC ORGANISM: MURINE  
CC FEATURE:  
CC NAME/KEY: CDS  
CC LOCATION: 1..549  
CC OTHER INFORMATION: /product= "MATURE mbMP3"  
SQ SEQUENCE 779 BP; 241 A; 164 C; 188 G; 186 T; 0 OTHER.

Query Match 1.1%; Score 27; DB 1; Length 779;  
Best Local Similarity 76.5%; Pred. No. 3.33e-02;  
Matches 39; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 726 TGAGATTAGTAAGTGTGATCTCTAAATAAAAAAAAAAAAAAAAAAAAAA 776  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Cp 1046 TGACATTTTAGAATGTGTCTTCTAAAGAGAAAAAATACACTAAAA 996

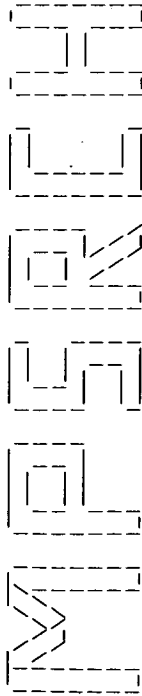
RESULT 15  
ID US-07-781-034-4 STANDARD; DNA; UNC: 1189 BP.  
AC xxxxxx  
DT  
DE Sequence 4, Application US/07781034  
CC Sequence 4, Application US/07781034  
CC Patent No. 5442050  
CC GENERAL INFORMATION:  
CC APPLICANT: Fishman, Jay A.  
CC TITLE OF INVENTION: Molecular Cloning of Antigens Shared By  
CC TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii  
CC NUMBER OF SEQUENCES: 4  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
CC STREET: Two Militia Drive  
CC CITY: Lexington  
CC STATE: Massachusetts  
CC COUNTRY: USA  
CC ZIP: 02173  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/781,034  
CC FILING DATE: 19911018  
CC CLASSIFICATION: 424  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Granahan, Patricia  
CC REGISTRATION NUMBER: 32,227  
CC REFERENCE/DOCKET NUMBER: MGH91-02A  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 861-6240  
CC TELEFAX: (617) 861-9540  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1189 base pairs  
CC TYPE: NUCLEIC ACID  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA to mRNA  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
SQ SEQUENCE 1189 BP; 370 A; 188 C; 228 G; 403 T; 0 OTHER.

Query Match 1.1%; Score 27; DB 1; Length 1189;

Best Local Similarity 84.6%; Pred. No. 3.33e-02;  
Matches 33; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 584 ATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 622  
||||| ||||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 164 ATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 202

Search completed: Fri Oct 22 03:25:11 1999  
Job time : 226 secs.

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Fri Oct 22 01:10:45 1999; MasPar time 4085.50 Seconds  
Tabular output not generated. 1400.561 Million cell updates/sec

Title: >US-09-040-485-1  
Description: (1-2442) from US09040485.seq  
Perfect Score: 2442  
N.A. Sequence: 1 CGCGAGCTTGAAGGACACAA.....GGTTACTTTAAATATTTT 2442  
Comp: GCCCTCGAAGCTTCCGTGTGT.....CCAATTGAAATTTATAAAAA

Scoring table:  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-est58  
1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2  
6:em\_est9 7:em\_gss1  
genbank-est11  
8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13  
13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33 33:gb\_est34  
34:gb\_est35 35:gb\_est36 36:gb\_est37 37:gb\_gss1 38:gb\_gss2  
39:gb\_gss3 40:gb\_gss4 41:gb\_gss5 42:gb\_gss6

Statistics: Mean 12.560; Variance 3.097; scale 4.056

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	497	20	4	851	9 AA203198	2x57a02.r1 Soares_feta 0.00e+00
2	488	19	2	502	32 N31186	Yx64f08.r1 Soares_mela 0.00e+00
3	434	17	8	600	22 AI084624	Ox78c04.x1 Soares_sene 0.00e+00
4	432	17	7	436	13 AA424475	Zv82e07.r1 Soares_tota 0.00e+00
5	400	16	4	424	35 AA037341	Zc52d12.r1 Soares_sene 0.00e+00
6	384	15	7	540	33 W03046	Za04d08.r1 Soares_mela 0.00e+00
7	367	15	0	585	9 AA176260	Zp29h01.s1 Stratagene 0.00e+00
8	363	14	9	488	19 T69872	Yx18e11.r1 Stratagene 0.00e+00
9	365	14	9	574	24 AI267311	acg3e03.x1 Stanley Fro 0.00e+00
10	350	14	3	575	34 W44451	Zc28d03.r1 Soares_sene 0.00e+00

C	11	344	14.1	598	32	N22713	yx54f08.s1 Soares mela	0.00e+00
12	340	13.9	350	8	T07611	EST05501 Fetal brain,	0.00e+00	
C	13	334	13.7	347	32	N26443	Yx23d03.r1 Soares mela	0.00e+00
14	327	13.4	598	33	N48349	YY79c12.s1 Soares_mult	0.00e+00	
C	15	325	13.3	392	32	D82780	HUMHBC2999 Human pancr	0.00e+00
16	325	13.3	392	8	T48492	hbc2999 Human pancreat	0.00e+00	
C	17	320	13.1	452	18	AA788915	ae93h04.s1 Stratagene	0.00e+00
18	321	13.1	471	15	AA602346	no89h06.s1 NCI_CGAP_AA	0.00e+00	
C	19	320	13.1	474	18	AA827544	of04d07.s1 NCI_CGAP_CO	0.00e+00
20	313	12.8	404	31	H26237	Y153h03.r1 Soares brea	0.00e+00	
C	21	311	12.7	480	28	AI498630	tm47a05.x1 NCI_CGAP_K1	0.00e+00
22	305	12.5	379	32	H81879	Yv83e11.r1 Soares mela	0.00e+00	
C	23	306	12.5	482	18	AA767510	oa90h03.s1 NCI_CGAP_GC	0.00e+00
24	303	12.4	549	9	AA176259	zp29h01.r1 Stratagene	0.00e+00	
C	25	289	11.8	315	34	C00261	HUMGS0005943 Human adu	0.00e+00
26	281	11.5	290	15	AA593699	nm64b10.s1 NCI_CGAP_Br	0.00e+00	
C	27	277	11.3	438	25	AI335659	qt31a03.x1 Soares_preg	0.00e+00
28	273	11.2	289	11	AA304742	EST175807 Bone VII Hom	0.00e+00	
C	29	274	11.2	426	9	AA157848	zo70h02.s1 Stratagene	0.00e+00
30	270	11.1	361	16	AA669808	ag36d03.s1 Jia bone ma	0.00e+00	
C	31	266	10.9	397	36	AA079718	zm20c10.s1 Stratagene	0.00e+00
32	265	10.9	417	15	AA576235	nm61f03.s1 NCI_CGAP_Br	0.00e+00	
C	33	263	10.8	279	32	H96029	Yv90f12.r1 Soares mela	0.00e+00
34	258	10.6	405	17	AA702708	zi90h09.s1 Soares_feta	0.00e+00	
C	35	258	10.6	446	9	AA159372	zo78h07.s1 Stratagene	0.00e+00
36	255	10.4	290	31	H25536	Y147g11.r1 Soares brea	0.00e+00	
C	37	250	10.2	399	25	AI307704	tb27h11.x1 NCI_CGAP_K1	0.00e+00
38	250	10.2	404	25	AI290103	qw26c10.x1 NCI_CGAP_Ut	0.00e+00	
C	39	248	10.2	410	14	AA541348	n177b06.s1 NCI_CGAP_Pr	0.00e+00
40	245	10.0	265	17	AA745282	nv48h06.s1 NCI_CGAP_Ew	0.00e+00	
C	41	244	10.0	394	28	AI540316	tz34e11.x1 NCI_CGAP_Ut	0.00e+00
42	242	9.9	412	32	H99385	Yx20f03.s1 Soares mela	0.00e+00	
C	43	237	9.7	391	32	N23024	Yx65f10.s1 Soares mela	0.00e+00
44	238	9.7	454	14	AA523182	n156g09.s1 NCI_CGAP_Pr	0.00e+00	
C	45	226	9.3	324	15	AA601990	no88h08.s1 NCI_CGAP_AA	0.00e+00

ALIGNMENTS

RESULT 1 AA203198 851 bp mRNA EST 24-JAN-1997  
LOCUS zx57a02.r1 Soares\_fetal\_liver\_spleen\_INFUS\_S1 Homo sapiens cdna  
DEFINITION clone IMAGE:446570 5', mRNA sequence.  
AA203198  
NID g1798908  
VERSION AA203198.1 GI:1798908  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and  
Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:693599.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amerham  
High quality sequence stop: 419.  
FEATURES  
Location/Qualifiers  
1..851  
Source

```

/organism="Homo sapiens"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subcloned version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AAGTGAAGAATTAAATAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="Soares_fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT      234 a 157 c 201 g 258 t 1 others
ORIGIN

```

```

Query Match      20.4%; Score 497; DB 9; Length 851;
Best Local Similarity 97.6%; Pred. No. 0.00e+00;
Matches 530; Conservative 0; Mismatches 9; Indels 4; Gaps 4;

Db 1 TTTTAGAAAACCTGTATAAATTACTGTGCATACTTAAGATTATTCTGCCTTGG-TA 59
QY 1647 TTTTAGAAAACCTGTATAAATTACTGTGCATACTTAAGATTATTCTGCCTTGGCTA 1706

Db 60 ATTGAGTAATCCCTCCAGCACTAGAGACCGCTCAGTGTCTTACTAGATGAACCTCAGT 119
QY 1707 ATTGAGTAATCCCTCCAGCACTAGAGACCGCTCAGTGTCTTACTAGATGAACCTCAGT 1766

Db 120 AACGCCCTGAGCTGGTGTGATTGAGGATGTGAAAAA-GCTCAGAGCCCGATGCCCTGC 178
QY 1767 AACGCCCTGAGCTGGTGTGATTGAGGATGTGAAAAAAGCTCAGAGCCCGATGCCCTGC 1826

Db 179 TGCATTATTCAGGCAATGAGCTTTTCTTCTACACTGAAGATTCTCTCTTAATTAAT 238
QY 1827 TGCATTATTCAGGCAATGAGCTTTTCTTCTACACTGAAGATTCTCTCTTAATTAAT 1886

Db 239 GTGCTTTATTTTGGGCTCAGAAATAATGCTCTGTGTAATAATATCTTTGTCAGAAAAG 298
QY 1887 GTGCTTTATTTTGGGCTCAGAAATAATGCTCTGTGTAATAATATCTTTGTCAGAAAAG 1946

Db 299 AAGGTAGTACCACATCATTTTGAAGAGCACCATGAGCAACTATAGCAAGCCATAAGAA 358
QY 1947 AAGGTAGTACCACATCATTTTGAAGAGCACCATGAGCAACTATAGCAAGCCATAAGAA 2006

Db 359 GTGCTTTGATCGATATATTAGGGGTAGCTCTTGATTTGTTAATTAAGATAAGGTGAC 418
QY 2007 GTGCTTTGATCGATATATTAGGGGTAGCTCTTGATTTGTTAATTAAGATAAGGTGAC 2066

Db 419 TTTTCCCTCC-TGCTGTAGGATTAATCAAGATCAAGTACTCTATATGTTTATTACTATAGA 477
QY 2067 TTTTCCCTCCCTTTTAGGATTAATCAAGATCAAGTACTCTATATGTTTATTACTATAGA 2126

Db 478 TCATAGTTATTATACATGTAGTGAAGTCTGCATGGGCTGCTGATGTTGATATGAAGCCT 537
QY 2127 TCATAGTTATTATACATGTAGTGAAGTCTGCATGGGCTGCTGATGTTGATATGAAGCCT 537

Db 538 GAA 540
QY 2186 GAA 2188

```

```

RESULT 2
LOCUS N31186 502 bp mRNA EST 10-JAN-1996
DEFINITION Yx64f08.r1 Soares melanocyte 2Nbhm Homo sapiens cDNA clone
IMAGE:266535 5', mRNA sequence.
ACCESSION N31186
NID g1153585
VERSION N31186.1 GI:1151585

```

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 502)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlffing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:693323.

```

```

Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
High quality sequence stops: 331
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: T7
High quality sequence stop: 331.
Location/Qualifiers

```

```

FEATURES
source

```

```

1..502
/organism="Homo sapiens"
/notes="Vector: pT7T3D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGGCGGAGTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."
/db_xref="GDB:387617"
/map="19"
/clone="IMAGE:266535"
/clone_lib="Soares melanocyte 2Nbhm"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT      165 a 69 c 77 g 190 t 1 others
ORIGIN

```

```

Query Match      19.2%; Score 468; DB 32; Length 502;
Best Local Similarity 98.2%; Pred. No. 0.00e+00;
Matches 491; Conservative 0; Mismatches 6; Indels 3; Gaps 3;

Db 1 GGGTTTTCATATATATTAGTAATTAACACTGTTGGACTGACCTGTGTACACTGTG 50
QY 1205 GGGTTTTCATATATATTAGTAATTAACACTGTTGGACTGACCTGTGTACACTGTG 1264

Db 61 TTAACATGATTTAAAGCTATTAAAGTACTTTGTGTAGCACTCTTAAACACCTAAC 120
QY 1265 TTAACATGATTTAAAGCTATTAAAGTACTTTGTGTAGCACTCTTAAACACCTAAC 1324

Db 121 AGAGATCATCTAGCTGTGAAGATTGAGTTGTATATACCTGCACCTGATATCTTATCA 180
QY 1325 AGAGATCATCTAGCTGTGAAGATTGAGTTGTATATACCTGCACCTGATATCTTATCA 1384

Db 181 AAAATTCTACATGCTTTAAAGTGTTCAGATTAACTTTTGAATTTTGTAGCTTTT 240
QY 1385 AAAATTCTACATGCTTTAAAGTGTTCAGATTAACTTTTGAATTTTGTAGCTTTT 1444

Db 241 AGCTGATTAATAGAAAATTTAATTTTCAGTGAAGCTTTTAATTTATCATTTTAT 300

```

QY 1445 AGCTGATTAATAGAAAAAT-ATAATTTTCAGTGAAGTTTTAAATATATCATTTATTTAT 1503  
 Db 301 TTTTAAATGAGAGGGAAGCTGAATCCCTTTGTTAAAGACACAGGAAAAAGATGGC 360  
 QY 1504 TTTTAAATGAGAGGGAAGCTGAATCCCTTTGTTAAAGACACAGGAAAAAGATGGC 1563  
 Db 361 CCTACTATTATCATGCAAAATGCTTTGTTGGCACCTCAGATTAATCATATAATAGCTAT 420  
 QY 1564 CCTACTATTATCATGCAAAATGCTTTGTTGGCACCTCAGATTAATCATATAATAGCTAT 1623  
 Db 421 AGTCTCTCCAGCATTTGTTAAATTTNAGAAACCTGTTATAAATTAAGTGTGTCATAACT 480  
 QY 1624 AGTCTCTTC-AGCATTTGTTAAATTTTAGAAAACCTGTTATAAATTAAGTGTGTCATAACT 1682  
 Db 481 TTTAAGATTATTCGCCCTT 500  
 QY 1683 TA-AGATTATTCGCCCTT 1701

RESULT 3  
 LOCUS AI084624 600 bp mRNA EST 01-OCT-1998  
 DEFINITION oz78c04.x1 Soares\_senescent\_fibroblasts\_NBHSF Homo sapiens cDNA  
 clone IMAGE:1681446 3', mRNA sequence.  
 ACCESSION AI084624  
 NID 93423047  
 VERSION AI084624.1 GI:3423047  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 600)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/hciogap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2282296.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1338 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 444.  
 Location/Qualifiers  
 1. .600

FEATURES  
 source  
 /organism="Homo sapiens"  
 /note="Vector: pT73D (Pharmacia) with a modified  
 polylinker V.TYPE: phagemid; Site\_1: Not I; Site\_2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dt)  
 primer [5']  
 TGTACCATCTCAAGTGGAGCGCGCGCATTTTTTTTTTTTTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M.Fatima Bonaldo.  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1681446"  
 /clone\_lib="Soares\_senescent\_fibroblasts\_NBHSF"  
 /tissue\_type="senescent fibroblast"  
 /lab\_host="DH10B (ampicillin resistant)"  
 BASE COUNT 207 a 106 c 74 g 213 t  
 ORIGIN  
 Query Match 17.8%; Score 434; DB 22; Length 600;  
 Best Local Similarity 99.18; Pred.No. 0.00e+00;  
 Matches 453; Conservative 0; Mismatches 1; Indels 3; Gaps 1;  
 Db 147 AAAAATATTTAAAGTTAAACACAGCATAATGAATCCCTCAACGCTCCAGAGTCTTACAAAAA 206

Cp 2442 AAAAATATTTAAAGTTAAACACAGCATAATCAATCCCTCAACGCTCCAGAGTCTTACAAAAA 2383  
 Db 207 TCCAGCAAAATCTACTTTTGGCTCATTCATCAGTTCTTATGTCTCAGTCTTAGTTTCCCTAAA 266  
 Cp 2382 TCCAGCAAAATCTACTTTTACTCATTCATCAGTTCTTATGTCTCAGTCTTAGTTTCCCTAAA 2323  
 Db 267 AAAATATGGCTTTATAAAAGTAGCTTCTATAATTCACAAAATGAAGAGTTTATTATATAA 326  
 Cp 2322 AAAATATGGCTTTATAAAAGTAGCTTCTATAATTCACAAAATGAAGAGTTTATTATATAA 2263  
 Db 327 TTTGAGTATCATCTCTGTATCAGCAGACAGCAGCTTTAGAAAATTTATGCTTTTCTTAT 386  
 Cp 2362 TTTGAGTATCATCTCTGTATCAGCAGACAGCAGCTTTAGAAAATTTATGCTTTTCTTAT 2203  
 Db 387 TATCTTATTTAT-TCAGGTTTCATTACATCGAGTACCATGAGGAGCTCAGTACAT 443  
 Cp 2202 TATCTTATTTATTTTTCAGGTTTCATTACATCGAGTACCATGAGGAGCTCAGTACAT 2143  
 Db 444 TGTATAAATCTATGATCTATAGTATGATAAAATATAGAGTATCTTTGATTTTATCTTA 503  
 Cp 2142 TGTATAAATCTATGATCTATAGTATGATAAAATATAGAGTATCTTTGATTTTATCTTA 2083  
 Db 504 AAAGCAGGGGAAAAAGTCACTTCTTCTTAATGTTTAAACAAAATCAAGAGCTACCCCTAAT 563  
 Cp 2082 AAAGCAGGGGAAAAAGTCACTTCTTCTTAATGTTTAAACAAAATCAAGAGCTACCCCTAAT 2023  
 Db 564 ATATCGATCAAAACACTTCTTTATGGCTTTGCTTATAG 600  
 Cp 2022 ATATCGATCAAAACACTTCTTTATGGCTTTGCTTATAG 1986  
 RESULT 4  
 LOCUS AA424475 436 bp mRNA EST 16-OCT-1997  
 DEFINITION zv82e07.r1 Soares\_total\_fetus\_Nb2HF8\_9w Homo sapiens cDNA clone  
 IMAGE:760164 5', mRNA sequence.  
 ACCESSION AA424475  
 NID 92103427  
 VERSION AA424475.1 GI:2103427  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 436)  
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,  
 Moore,B., Schellenberg,K., Steptoe,W., Tan,F., Theising,B.,  
 White,Y., Wyllie,T., Waterston,R. and Wilson,R.  
 WashU-Merck EST Project 1997  
 Unpublished (1997)  
 COMMENT On Sep 21, 1992 this sequence version replaced gi:279345.

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 425.  
 Location/Qualifiers  
 1. .436  
 /organism="Homo sapiens"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from pooled 8-9 week  
 (total) fetus material with a Not I - oligo(dt) primer [5']  
 TGTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3']  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

/db\_xref="taxon:9606"

/clone\_lib="Soares:760164"

/dev\_stage="8-9 weeks"

/lab\_host="DH10B"

147 a 57 c 70 g 162 t

BASE COUNT  
ORIGIN

Query Match 17.7%; Score 432; DB 13; Length 436;

Best Local Similarity 99.5%; Pred. No. 0.00e+00;

Matches 434; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 GTGGAGTACTGTGTACACTGTGTTAAACATGATTTAAAGCTATTAAAGTACTTTG 60

QY 1240 GTGGAGTACTGTGTACACTGTGTTAAACATGATTTAAAGCTATTAAAGTACTTTG 1299

Db 61 TGTAGACACTCTTAAACAGCTACAGAGATCATATTAGCTGTGAAGATTGAGTTGTA 120

QY 1300 TGTAGACACTCTTAAACAGCTACAGAGATCATATTAGCTGTGAAGATTGAGTTGTA 1359

Db 121 TATACCTGCACGTATCTTATCAAAAATTTCTACATAGCTTTAAAGTGTTCAGATTAA 180

QY 1360 TATACCTGCACGTATCTTATCAAAAATTTCTACATAGCTTTAAAGTGTTCAGATTAA 1419

Db 181 CACATTTGAAATTTTGTAGCTTTTGTAGCTATTAATAGAAAAATTAATTTTCAAGTAA 240

QY 1420 CACATTTGAAACCTTTGTAGCTTTTGTAGCTATTAATAGAAAAATTAATTTTCAAGTAA 1479

Db 241 AGTTTAAATATCATTTATTTATTTTAAATGAGAGGGAAGCTGAATTCCTTGT 300

QY 1480 AGTTTAAATATCATTTATTTATTTTAAATGAGAGGGAAGCTGAATTCCTTGT 1539

Db 301 TAAGACACAGAAAGAAAGTGGCCCTACTATTATCATGAAAAATGCTTTGTGGCACC 360

QY 1540 TAAGACACAGAAAGAAAGTGGCCCTACTATTATCATGAAAAATGCTTTGTGGCACC 1599

Db 361 TCAGATTAAATCATATATAGCTATAGTCTCTTCAGCATTTGTTTAAATTTAGAAAACT 420

QY 1600 TCAGATTAAATCATATATAGCTATAGTCTCTTCAGCATTTGTTTAAATTTAGAAAACT 1659

Db 421 GTATAAATTTACTGGTG 436

QY 1660 GTATAAATTTACTGGTG 1675

RESULT 5 AA037341 424 bp mRNA EST 25-NOV-1996

LOCUS zc52d12.r1 Soares\_senescent\_fibroblasts\_NBHSF Homo sapiens cDNA

DEFINITION clone IMAGE:325943 5', mRNA sequence.

ACCESSION AA037341

NID g1512441

VERSION AA037341.1 GI:1512441

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 424)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlff, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

On Oct 18, 1995 this sequence version replaced gi:1023479.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 815 Std Error: 0.00

Seq Primer: -28M13 rev2 from Amersham

High quality sequence stop: 194.

Location/Qualifiers

1. 424

/organism="Homo sapiens"

/note="vector: pT73D (Pharmacia) with a modified

polylinker V.TYPE: phagemid; Site.1: Not I; Site.2: Eco

RI; 1st strand cDNA was primed with a Not I - oligo(dT)

primer [5'

TGTTACCAATCTGAAGTGGGAGCGCGCATTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M. Fatima Bonaldo.

/db\_xref="GDB:1257456"

/db\_xref="taxon:9606"

/clone="IMAGE:325943"

/clone\_lib="Soares\_senescent\_fibroblasts\_NBHSF"

/tissue\_type="senescent fibroblast"

/lab\_host="DH10B (ampicillin resistant)"

BASE COUNT 121 a 80 c 82 g 139 t 2 others

ORIGIN

Query Match 16.4%; Score 400; DB 35; Length 424;

Best Local Similarity 98.6%; Pred. No. 0.00e+00;

Matches 419; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

Db 1 GTAATTCCTCCAGCCTAGACCGCTCAGTCTCTTACTAGTAACTCAGTAACGC 60

QY 1712 GTAATTCCTCCAGCCTAGACCGCTCAGTCTCTTACTAGTAACTCAGTAACGC 1771

Db 61 CTGAGCTGGTGTGATGAGGATGTGTGAAA-GCTCACAGAGCCCGATGCTGCTA 119

QY 1772 CTGAGCTGGTGTGATGAGGATGTGTGAAAAGCTCACAGAGCCCGATGCTGCTA 1831

Db 120 TTTACGCAATGAGCCTTTCTTTCTACACTGAAGATTTCTTCTTATTAATGTGT 179

QY 1832 TTTACGCAATGAGCCTTTCTTTCTACACTGAAGATTTCTTCTTATTAATGTGT 1891

Db 180 TTTTGTGGCTCAGAAATAATGCTCTGTGAAATAATCTTTGTGAGAAAGAAAGT 239

QY 1892 TTTTGTGGCTCAGAAATAATGCTCTGTGAAATAATCTTTGTGAGAAAGAAAGT 1951

Db 240 AGTACACATCATTTTGAAGGACCATGAGCACTATAAGCAAGCCATAGAGTGTGT 299

QY 1952 AGTACACATCATTTTGAAGGACCATGAGCACTATAAGCAAGCCATAGAGTGTGT 2011

Db 300 TTGATCGATATATTAGGGGTAGCTCTTGGTTTGTACATTAAGTAAAGTGACNTTT 359

QY 2012 TTGATCGATATATTAGGGGTAGCTCTTGGTTTGTACATTAAGTAAAGTGACNTTT 2071

Db 360 CCCCTGCTTTTAGATTAAAAATCNAAGATCTCCCATATTTTATCATTATAGATCA 419

QY 2072 CCCCTGCTTTTAGATTAAAAATCNAAGATCTCCCATATTTTATCATTATAGATCA 2129

Db 420 TAGTT 424

QY 2130 TAGTT 2134

RESULT 6

LOCUS W03046

DEFINITION za04d08.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone

IMAGE:291567 5', mRNA sequence.

ACCESSION W03046

18-APR-1996

EST

mRNA

540 bp



```

91275093
VERSION W03046.1 GI:1275093
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 540)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 215.

FEATURES
source
1..540
/organism="Homo sapiens"
/Note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGGGACGGCGCGAGTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."
/db_xref="GDB:3884697"
/db_xref="taxon:9606"
/clone="IMAGE:291567"
/clone_lib="Soares melanocyte 2NbHM"
/sex="Male"
/tissue="type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 157 a 93 c 124 g 163 t
ORIGIN
1 TAAATGAGAGGGGAAAGCTGAATTCGGTCTTAAGACACAGGAAAGAAAGATGCCCTAC 60
|||||
QY 1509 TAAATGAGAGGGGAAAGCTGAATTCGGTCTTAAGACACAGGAAAGAAAGATGCCCTAC 1568
|||||
Db 61 TATTATCATGCAAAATGCTTTGTTGGCACCTCAGATTATCATATATATAGCTATAGTCT 120
|||||
QY 1569 TATTATCATGCAAAATGCTTTGTTGGCACCTCAGATTATCATATATATAGCTATAGTCT 1628
|||||
Db 121 CTTGACATTTGTTTAAATTTTGAAGAACTGTATAAATTTACTGGTGCAATTAAGA 180
|||||
QY 1629 CTTGACATTTGTTTAAATTTTGAAGAACTGTATAAATTTACTGGTGCAATTAAGA 1688
|||||
Db 181 TTATTCGCTTTGGG-TAATGAGTAAATCCCTCCAGCACTAGACCCCTCAGTGCTC 239
|||||
QY 1689 TTATTCGCTTTGGGTAATGAGTAAATCCCTCCAGCACTAGACCCCTCAGTGCTC 1748
|||||
Db 240 TTACTAGTAGAAGTCAAGTCAAGCTTGGCTGGTGTGATGAGGATCTGTGAAGAA-GCTC 298
QY 1749 TTACTAGTAGAAGTCAAGTCAAGCTTGGCTGGTGTGATGAGGATCTGTGAAGAAAGCTC 1808
|||||

Query Match 15.7%; Score 384; DB 33; Length 540;
Best Local Similarity 95.0%; Pred. No. 0.00e+00;
Matches 514; Conservative 0; Mismatches 6; Indels 21; Gaps 20;

Db 1 TAAATGAGAGGGGAAAGCTGAATTCGGTCTTAAGACACAGGAAAGAAAGATGCCCTAC 60
|||||
QY 1509 TAAATGAGAGGGGAAAGCTGAATTCGGTCTTAAGACACAGGAAAGAAAGATGCCCTAC 1568
|||||
Db 61 TATTATCATGCAAAATGCTTTGTTGGCACCTCAGATTATCATATATATAGCTATAGTCT 120
|||||
QY 1569 TATTATCATGCAAAATGCTTTGTTGGCACCTCAGATTATCATATATATAGCTATAGTCT 1628
|||||
Db 121 CTTGACATTTGTTTAAATTTTGAAGAACTGTATAAATTTACTGGTGCAATTAAGA 180
|||||
QY 1629 CTTGACATTTGTTTAAATTTTGAAGAACTGTATAAATTTACTGGTGCAATTAAGA 1688
|||||
Db 181 TTATTCGCTTTGGG-TAATGAGTAAATCCCTCCAGCACTAGACCCCTCAGTGCTC 239
|||||
QY 1689 TTATTCGCTTTGGGTAATGAGTAAATCCCTCCAGCACTAGACCCCTCAGTGCTC 1748
|||||
Db 240 TTACTAGTAGAAGTCAAGTCAAGCTTGGCTGGTGTGATGAGGATCTGTGAAGAA-GCTC 298
QY 1749 TTACTAGTAGAAGTCAAGTCAAGCTTGGCTGGTGTGATGAGGATCTGTGAAGAAAGCTC 1808
|||||

Query Match 15.0%; Score 367; DB 9; Length 585;
Best Local Similarity 93.3%; Pred. No. 0.00e+00;
Matches 420; Conservative 0; Mismatches 24; Indels 6; Gaps 4;

```

Db 137 AAAAATATTTAAAGTTAACACAGCATAATGAATCCTCAACGTCAGAGTTCTCAAAAA 196  
|||||  
Cp 2442 AAAAATATTTAAAGTTAACACAGCATAATGAATCCTCAACGTCAGAGTTCTCAAAAA 2383  
|||||  
Db 197 TCCAGCAAACTTACTTTTACTCATTCATCATAGTTCTATGTCTCACTCCTTAGTTTCCCTAAA 256  
|||||  
Cp 2382 TCCAGCAAACTTACTTTTACTCATTCATCATAGTTCTATGTCTCACTCCTTAGTTTCCCTAAA 2323  
|||||  
Db 257 AAAATATGGCTTTATAAAGTAGTCTTATAATTCACAAATGAAGAGTTTATATATA 316  
|||||  
Cp 2322 AAAATATGGCTTTATAAAGTAGTCTTATAATTCACAAATGAAGAGTTTATATATA 2263  
|||||  
Db 317 TTTGAGTATCATCTGTATCACCGACAGCAGCTTTAGAAAATATATGCTTTCTTAT 376  
|||||  
Cp 2262 TTTGAGTATCATCTGTATCACCGACAGCAGCTTTAGAAAATATATGCTTTCTTAT 2203  
|||||  
Db 377 TATCTTATTTATTTTCAGTTTTCATACATCGAGTACCCATGCCAGGACTCACTACA 433  
|||||  
Cp 2202 TATCTTATTTATTTTCAGTTTTCATACATCGAGTACCCATGCCAGGACTCACTACA 2144  
|||||  
Db 434 TTGNATAAATACTATGATCTATAGTGTATAGTGTATAGTGTATAGTGTATAGTGTAT 493  
|||||  
Cp 2143 TTGTATAAATACTATGATCTATAGTGTATAGTGTATAGTGTATAGTGTATAGTGTAT 2084  
|||||  
Db 494 AAAAGCAGGGGGAAGGCCCCCTATCTTATGTAACCAATCCAGGGGCTACCCCA 553  
|||||  
Cp 2083 AAAAGCAGGGGGAAGGCCCCCTATCTTATGTAACCAATCCAGG-CTACCCCTA 2025  
|||||  
Db 554 ATATTGGGTCAA-CCNCTTCTCTATGNTT 582  
|||||  
Cp 2024 ATATATCATCAACCACTTCTTATGGCTT 1995  
|||||

RESULT 8  
LOCUS T69872 468 bp mRNA EST 23-FEB-1995  
DEFINITION Yc18ell.r1 Stratagene lung (#937210) Homo sapiens cDNA clone  
IMAGE:81068 5', mRNA sequence.  
ACCESSION T69872  
NID 9681020  
VERSION T69872.1 GI:681020  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 468)  
AUTHORS Holman,M., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Hillier,L., Kucaba,T., Lacy,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,  
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.  
WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 504  
High quality sequence stops: 338 Source: IMAGE Consortium, LLNL This  
clone is available royalty-free through LLNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.  
Insert Length: 504 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 338.  
Location/Qualifiers  
1..468  
/organism="Homo sapiens"  
/note="Organ: lung; Vector: pBluescript SK-; Site:1;  
EcoRI; Site2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT, normal lung. Average insert size: 1.0 Kb;

Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTITTTTTTTTTT 3'  
/db\_xref="GDB:484685"  
/db\_xref="taxon:9606"  
/clone\_lib="Stratagene lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"  
/lab\_host="SOLR cells (kanamycin resistant)"  
BASE COUNT 159 a 90 c 114 g 102 t 3 others  
ORIGIN

Query Match 14.9%; Score 363; DB 19; Length 468;  
Best Local Similarity 94.7%; Pred. No. 0.00e+00;  
Matches 430; Conservative 0; Mismatches 15; Indels 9; Gaps 9;  
Db 6 GAGGATAGATTGAGACCTCGAACCTGAAGTATCTCATGAAGAAACCGAGCATAGTTAC 65  
|||||  
Qy 472 GATGATAGATTGAGACCTCGAACCTGAAGTATCTCATGAAGAAACCGAGCATAGTTAC 531  
|||||  
Db 66 CACGTGGAAGACAGAGTTTTCACAGACTGTATATCAGGATATGGAAGAGATGATGCTGAG 125  
|||||  
Qy 532 CACGTGGAAGACAGAGTTTTCACAGACTGTATATCAGGATATGGAAGAGATGATGCTGAG 591  
|||||  
Db 126 CAGGAATCCAGATTCAGTGAACCTAGTAGAGATGAAGATTGCACCATGATACA 185  
|||||  
Qy 592 CAGGAATCCAGATTCAGTGAACCTAGTAGAGATGAAGATTGCACCATGATACA 651  
|||||  
Db 186 GATGATGAACATACCAAGTCTATGAGGAACAGCAGTATATGAACCTCTAGAAATGAA 245  
|||||  
Qy 652 GATGATGAACATACCAAGTCTATGAGGAACAGCAGTATATGAACCTCTAGAAATGAA 711  
|||||  
Db 246 GGGATAGAAATACAGAAAGTAACTGCTCCCTTGGAGATATCTCTAGAAAGATTACA 305  
|||||  
Qy 712 GGGATAGAAATACAGAAAGTAACTGCTCCCTTGGAGATATCTCTAGAAAGATTACA 770  
|||||  
Db 306 GGTATTTAGGAAGAGTAAAGCATTTTCTGTTGGGAGGACACGAGGAGTACCACCA 365  
|||||  
Qy 771 GGTATTTAG-AGAAGTAAAGCATTTTCTGTTGGAG- AACACGAGGAGTACCACCA 828  
|||||  
Db 366 GNTACTTTAAAGCTTCCAAAAGATTGCCCCCTACACACAGGCGGCGCCGCTTAA 425  
|||||  
Qy 829 GATACCTT-AAAGCTTC-AAAGAGACTGCCCTTACCACACAGG-AGG-ACCAGCT-AA 882  
|||||

RESULT 9  
LOCUS A1267311 574 bp mRNA EST 17-NOV-1998  
DEFINITION ag63e03.x1 Stanley Frontal SN pool 2 Homo sapiens cDNA clone  
IMAGE:2035612 similar to SW:ASPH\_HUMAN Q12797 ASPARTYL/ASPARAGINYL  
BETA-HYDROXYLASE ; mRNA sequence.  
ACCESSION A1267311  
NID 93886478  
VERSION A1267311.1 GI:3886478  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 574)  
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2150666.  
Contact: Wilson RK  
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 439.

# FEATURES

Location/Qualifiers  
 1..574  
 /organism="Homo sapiens"  
 /note="Organ: Brain; Vector: pCR2.1-TOPO (Invitrogen);  
 Site 1: EcoRI; Total RNA (purified with Trizol and DNaseI  
 before use) was reverse transcribed using a modified  
 oligo-dT primer containing RsaI and HindIII sites.  
 Double-stranded cDNA was digested with RsaI, resulting in  
 blunt ended cDNA of an average 0.1-2 kb in length.  
 Digested cDNA was split into two sets, one used as is as  
 the driver, the other set was split in half again and each  
 half linked to a different adaptor  
 (5'-TCGACGCCGCCGCCGCCAGGT-3' or 5'-  
 AGGCGTGTGGGAGGCGGT-3'), to be used as tester.  
 Subtraction was performed using the Clontech PCR Select  
 cDNA subtraction kit. Pool of two schizophrenics, male age  
 44 and female age 56 (S-116, S-118) subtracted by pool of  
 two mentally normal male individuals ages 41 and 53  
 (S-124, S-141). Tissues were obtained from the Stanley  
 Neuropathology Consortium (www.stanleylab.org). Library  
 constructed and subtracted by Dr. Nancy Johnston [(410)  
 614-3918, nlj@welchlink.welch.jhu.edu].  
 /db\_xref="taxon:9606"  
 /map="21q"  
 /clone="IMAGE:2035612"  
 /clone\_lib="Stanley Frontal SN pool 2"  
 /tissue\_type="frontal lobe (see description)"  
 /lab\_host="DH10B (phage-resistant)"  
 164 a 150 c 109 g 150 t 1 others

Query Match 14.9%; Score 365; DB 24; Length 574;  
 Best Local Similarity 99.5%; Pred. No. 0.00e+00;  
 Matches 367; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 18 GTACATTTTCATGAGAAGGACTGAATTTGTTCTTTGCTTCATCTCGATATTCGTGG 77  
 |||||  
 Cp 373 GTACCATTTTCATGAGAAGGACTGAATTTGTTCTTTGCTTCATCTCGATATTCGTGG 314  
 |||||

Db 78 GTTCTCGCTCCACAGGAACCTGCTCTCGGCTCAGTGTGGTCAAGCCTCTTCTGGCG 137  
 |||||  
 Cp 313 GTTCTCGCTCCACAGGAACCTGCTCTCGGCTCAGTGTGGTCAAGCCTCTTCTGGCG 254  
 |||||

Db 138 GGACTGCTGGCTCTGAAGTAGATCTCTTTTAAGTCTCTATAAACTTTCGCATCATCCA 197  
 |||||  
 Cp 253 GGACTGCTGGCTCTGAAGTAGATCTCTTTTAAGTCTCTATAAACTTTCGCATCATCCA 194  
 |||||

Db 198 CATCAAAATCTCCATCACCATCAGCATCATAGATTCCTAGTTTCTCCPAGAACTTCCCTCAT 257  
 |||||  
 Cp 193 CATCAAAATCTCCATCACCATCAGCATCATAGATTCCTAGTTTCTCCPAGAACTTCCCTCAT 134  
 |||||

Db 258 AGTCAACAGATCAAAACCAACGACAGCTACAGATGTCACAGCGCCAGCAATCA 317  
 |||||  
 Cp 133 AGTCAACAGATCAAAACCAACGACAGCTACAGATGTCACAGCGCCAGCAATCA 74  
 |||||

Db 318 CCATAACACAGTGAAGTGAAGTTCCTGAGAGTCGGCTTCTCCCTCCATTCCTGTGTC 377  
 |||||  
 Cp 73 CCATAACACAGTGAAGTGAAGTTCCTGAGAGTCGGCTTCTCCCTCCATTCCTGTGTC 14  
 |||||

Db 378 CTCATGCT 386  
 |||||  
 Cp 13 CTCATGCT 5

RESULT 10

# LOCUS

DEFINITION zc28b03.r1 Soares senescent fibroblasts\_NbHSF Homo sapiens CDNA

# ACCESSION

NID W44451  
 VERSION g1329952  
 KEYWORDS W44451.1 GI:1329952

# SOURCE

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

# REFERENCE

AUTHORS Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,  
 Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,  
 Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,  
 Trevasaki E., Waterston R., Williamson A., Woldmann P. and  
 Wilson R.

# TITLE

JOURNAL The WashU-Merck EST Project  
 COMMENT Unpublished (1995)

# COMMENT

On May 8, 1995 this sequence version replaced gi:801247.  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

# FEATURES

Location/Qualifiers  
 1..575  
 /organism="Homo sapiens"  
 /note="Vector: pT7T3D (Pharmacia) with a modified  
 polylinker V\_LYFE; phagemid; Site\_1: Not I; Site\_2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer [5'  
 TGTACCAATCTGAAGTGGGCGGCGCATTTTTTTTTTTTTTTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT7T3 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M. Fatima Bonaldo.  
 /db\_xref="GDB:1255109"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:323597"  
 /clone\_lib="Soares senescent fibroblasts\_NbHSF"  
 /tissue\_type="senescent fibroblast"  
 /lab\_host="DH10B (ampicillin resistant)"  
 171 a 94 c 118 g 182 t 10 others

# BASE COUNT

ORIGIN  
 Query Match 14.3%; Score 350; DB 34; Length 575;  
 Best Local Similarity 96.9%; Pred. No. 0.00e+00;  
 Matches 406; Conservative 0; Mismatches 4; Indels 9; Gaps 9;

Db 1 ACTTTGTTAGCACTCTTAAAAACGCTAACAGAGATCANCATTCAGTGTGAAGATTGA 60  
 |||||  
 QY 1294 ACTTTGTTAGCACTCTTAAAAACGCTAACAGAGATCANCATTCAGTGTGAAGATTGA 1353  
 |||||

Db 61 GTTGTATATACCTGCACATGATATCTTATCAAAATTTCTACATTCAGTTAAGTTCA 120  
 |||||  
 QY 1354 GTTGTATATACCTGCACATGATATCTTATCAAAATTTCTACATTCAGTTAAGTTCA 1413  
 |||||

Db 121 GATTAAACACATTTTGAATTTTGTAGCTTTTAGCTGATTAATAAGAAATTAATATTC 180  
 |||||  
 QY 1414 GATTAAACACATTTTGAATTTTGTAGCTTTTAGCTGATTAATAAGAAATTAATATTC 1473  
 |||||

Db 181 AGTGAAGTTTAAATTTATCATTTTATTTATTTTAAATGAGAGGGGAAAGCTGAATTT 240  
 |||||  
 QY 1474 AGTGAAGTTTAAATTTATCATTTTATTTATTTTAAATGAGAGGGGAAAGCTGAATTT 1533  
 |||||



the Not I and Eco RI sites of a modified pTT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."									
/db_xref="GDB:387207"									
/db_xref="taxon:9606"									
/clone="IMAGE:262565"									
/clone_lib="Soares melanocyte 2NBHM"									
/sex="Male"									
/tissue_type="melanocyte"									
/lab_host="DH10B (ampicillin resistant)"									
BASE COUNT	101 a	45 c	64 g	135 t	2 others				
ORIGIN									
Query Match 13.7%; Score 334; DB 32; Length 347;									
Best Local Similarity 98.6%; Pred. No. 0.00e+00;									
Matches 342; Conservative 0; Mismatches 4; Indels 1; Gaps									
Db	1	GTGAGACAGCTTTAGTCTTNCCTGAATTTATGTGTGTTTTCCGACAGCTGGTTAATAAT	60						
Qy	1112	GTGAGACAGCTTTAGTCTTACCTGAATTTATGTGTGTTTTCCGACAGCTGGTTAATAAT	1171						
Db	61	TATATTGGTGATGAGCAGCAATGTGTGGCAGGGTTTCATATATTATTAGTAATTAA	120						
Qy	1172	TATATTGGTGATGAGCAGCAATGTGTGGCAGGGTTTCATATATTATTAGTAATTAA	1231						
Db	121	CACTAACCTGTTGGAGCTGACTTGTGTACACTGTGTTAAACATCATTTAAAGCTATTAAAGA	180						
Qy	1232	CACTAACCTGTTGGAGCTGACTTGTGTACACTGTGTTAAACATCATTTAAAGCTATTAAAGA	1291						
Db	181	GTACTTTTGTGTAGCACCTCTTAAACGCTTACAGAGATCATCATTAGCTGTGAAGATTT	240						
Qy	1292	GTACTTTTGTGTAGCACCTCTTAAACGCTTACAGAGATCATCATTAGCTGTGAAGATTT	1351						
Db	241	GAGTTGTATATACCTGCACTGATATCTTATCAAAAATTTCTACATTAGCTTTAAAGTGT	300						
Qy	1352	GAGTTGTATATACCTGCACTGATATCTTATCAAAAATTTCTACATTAGCTTTAAAGTGT	1411						
Db	301	CAGATTACACCTTTTGAATTTTGTAGCTTTNAGCTGGATTAATA	347						
Qy	1412	CAGATTACACCTTTTGAACCTTTTGTAGCTTTTAGCT-ATTAATA	1457						
RESULT 14									
LOCUS	N48349	598 bp	mrna	EST	14-FEB-1996				
DEFINITION	Y79C12.s1 Soares_multiple_sclerosis_2NBHMP Homo sapiens cdna clone IMAGE:279766 3', mRNA sequence.								
ACCESSION	N48349								
NID	N48349.1	GI:1189515							
VERSION	EST.								
KEYWORDS	human.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
AUTHORS	1 (bases 1 to 598) Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaaba,T., Le.M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.								
TITLE	The WashU-Merck EST Project								
JOURNAL	Unpublished (1995)								
COMMENT	On Apr 14, 1993 this sequence version replaced qi:785400.								

On Apr 18, 1993 this sequence version replaced g17033900.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Seq primer: ml3 -40 forward  
High quality sequence stop: 339.

FEATURES  
source  
1..598  
Location/Qualifiers  
/organism="Homo sapiens"  
/note="vector: p7T3D (Pharmacia) with a modified polylinker V.TYPE: phagemid; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5',  
TGTACCAATCTGAAGTGGGAGCGCGCATTTTTTTTTTTTTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis lesions from one patient was kindly provided by Dr. Kevin G. Becker (NINDS/NIH).  
/db\_xref="GDB:3898142"  
/db\_xref="taxon:9606"  
/map="3"  
/clone="IMAGE:279766"  
/clone\_lib="Soares\_multiple\_sclerosis\_2NBHMS"  
/sex="male"  
/tissue\_type="multiple sclerosis lesions"  
/dev\_stage="Age 46"  
/lab\_host="DH10B (ampicillin resistant)"  
BASE COUNT 192 a 120 c 92 g 186 t 8 others  
ORIGIN

Query Match 13.4%; Score 327; DB 33; Length 598;  
Best Local Similarity 89.6%; Pred. No. 0.00e+00;  
Matches 412; Conservative 0; Mismatches 39; Indels 9; Gaps 9;

Db 121 AAAAATATTTAAAGTAAACACAGCATAATGAATCCTCAACGTCAGAGTTCTCAAAAA 180  
Cp 2442 AAAAATATTTAAAGTAAACACAGCATAATGAATCCTCAACGTCAGAGTTCTCAAAAA 2383  
Db 181 TCCAGCAAACTTACTTTTGTCTCATTCATGCTTCTATGTCACCTCCTTAGTTTCCCTAAA 240  
Cp 2382 TCCAGCAAACTTACTTTTGTCTCATTCATGCTTCTATGTCACCTCCTTAGTTTCCCTAAA 2323  
Db 241 AAAATATGCTTTTAAAGTAGCTTCTATAATTCACAAAATGAAGAGTTTATATATA 300  
Cp 2322 AAAATATGCTTTTAAAGTAGCTTCTATAATTCACAAAATGAAGAGTTTATATATA 2263  
Db 301 TTTGAGTATCATCTCTGTATCACCAGCAGCAGCTTTAGAAAATTTATTGCTTT-TCC 359  
Cp 2262 TTTGAGTATCATCTCTGTATCACCAG-ACAGCACAGCTTTAGAAAATTTATTGCTTTCTTA 2204  
Db 360 CTATTATCNTATTTATTTCAGGNTTCATTACACATCGGAGTACCCATGGCAGACTCACTA 419  
Cp 2203 TTATCTTATTATTATTTCAGGTTTCTATTACATCG-AGTACCCATG-CAGGACTCACTA 2146  
Db 420 CATTTGGTATATACTATGGATCATGATGGATATAAATATAGAGTATCTTGGATTTT 479  
Cp 2145 CATTTG-TATAATACTATG-ATCTATATG-ATAAATAATAGAGTATCTTTG-ATTTT 2090  
Db 480 AATCCCTAAAGCNGGGGAAAGGTCNCCCTATCTTATGTTTACCAAAATCCAGNGCTA 539  
Cp 2089 ATCTCTAAAGCAGGGGAAAGG-TCACTTATCTTATGTTTACAAAATCAAGAGCTA 2031  
Db 540 CCCCCTNATATCGGGTCAACCACTTCTCTANGCGTTGGCT 579  
Cp 2030 CCCCTAATATCGATCAACCACTTCTTATGCTTTGCT 1991

RESULT 15  
LOCUS D82780 392 bp mRNA EST 09-FEB-1996  
DEFINITION HUMHBC2999 Human pancreatic islet Homo sapiens cDNA, mRNA sequence.  
ACCESSION D82780  
NID 91183278  
VERSION D82780.1 GI:1183278

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Takeda, J.  
TITLE Human pancreatic islet ESTs  
JOURNAL Unpublished (1995)  
COMMENT  
Contact: Jun Takeda  
Institute for Molecular and Cellular Regulation, Gunma University  
3-39-15 Shova-machi, Maebashi Gunma 371, Japan  
Tel: 272-20-8856  
Fax: 272-20-8896  
Email: jtakeda@b.gunma-u.ac.jp.  
Location/Qualifiers  
source 1..392  
/organism="Homo sapiens"  
/note="vector: Lambda ZAPII; Site\_1: Eco RI; Site\_2: Xho I; mRNA was prepared from normal adult human islets. cDNA was directionally synthesized from the Xho I in the vector to the EcoRI site. cDNA was size fractionated to remove sequences <1000 bp in size."  
/db\_xref="taxon:9606"  
/clone\_lib="Human pancreatic islet"  
BASE COUNT 122 a 63 c 65 g 135 t 7 others  
ORIGIN

Query Match 13.3%; Score 325; DB 32; Length 392;  
Best Local Similarity 94.4%; Pred. No. 0.00e+00;  
Matches 370; Conservative 0; Mismatches 16; Indels 6; Gaps 6;

Db 1 AGCTTTAAGTGTTCAGATTAACACATTTTGGAAATTTTGTAGCTTTTAGCTGATTATTA 60  
Qy 1399 AGCTTTAAGTGTTCAGATTAACACATTTTGG-AAAACCTTTTGTAGCTTTTAGCTGATTATTA 1457  
Db 61 GAAAAATTAATTTTCAGTGAAGTTTTAAATTTATCATTTATTTTANATGAGA 120  
Qy 1458 GAAAAATTAATTTTCAGTGAAGTTTTAAATTTATCATTTATTTTAAATGAGA 1517  
Db 121 GGGGAAAGCTGAAATTCCTTGTTAAGACACAGGAAAGAAATGCGCTTACTATTATCAT 180  
Qy 1518 GGGGAAAGCTGAAATTCCTTGTTAAGACACAGGAAAGAAATGCGCTTACTATTATCAT 1577  
Db 181 GCAAAATGCTTTGTGGCACCCTCAGATTAATCATATATAGCTATAGNCTCTTCAGCAT 240  
Qy 1578 GCAAAATGCTTTGTGGCACCCTCAGATTAATCATATATAGCTATAGTCTCTTCAGCAT 1637  
Db 241 TTGTTTAAATTTTAGAAAACCTGTATNAATTTACTGGTGCATAACTTAAGATTATTTCTGC 300  
Qy 1638 TTGTTTAAATTTTAGAAAACCTGTATNAATTTACTGGTGCATAACTTAAGATTATTTCTGC 1697  
Db 301 CTTTGGCTANTTTAGTAAATTCCTCCNCCAGCAGCTAGGAGACCTCCAAAGGGCTCTTACTA 360  
Qy 1698 CTTTGGCTANTTTAGTAAATTTCTCC-TCAGCAGCTAG-AGACCGCTCA-GTGCTCTTACTA 1754  
Db 361 GATTGAACCTCAAGNAACGCTTGGAGCTGGCT 392  
Qy 1755 GAT-GAATCA-GTAACGCTTGGAGCTGGCT 1784

Search completed: Fri Oct 22 03:08:45 1999  
Job time : 7080 secs.

\*\*\*\*\*  
W P S R F L  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:20:56 1999; MasPar time 11.52 Seconds  
Tabular output not generated. 470.550 Million cell updates/sec

Title: >US-09-040-485-2  
Description: (1-255) from US09040485.pep  
Perfect Score: 1758  
Sequence: 1 MVIALLGVTSAVVWFGLV.....IVEEVSIFPVEEQEVPPDT 255

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part2 2:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 31.856; Variance 172.331; scale 0.185

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length DB ID	Description	Pred. No.	
1	208	11.8	1018 19 R98747	P. vivax ESP-1 blood	2.20e-06	
2	191	10.9	562 14 R70491	Leucocytozoan protozo	3.40e-05	
3	190	10.8	412 19 W03626	Human thyrotropin GPR	3.99e-05	
4	186	10.6	1073 3 P60569	Sequence of the Ring	7.55e-05	
5	181	10.3	1132 17 R97866	Chicken leucocytozoan	1.67e-04	
6	178	10.2	332 14 R70492	Leucocytozoan protozo	2.30e-04	
7	179	10.1	360 19 W03627	Human follicle stimu	2.69e-04	
8	174	9.9	1529 18 R97985	CORK potassium channe	5.06e-04	
9	173	9.8	195 14 R70493	Leucocytozoan protozo	5.93e-04	
10	169	9.6	783 3 R05804	C-terminal of native	1.11e-03	
11	145	8.2	465 33 W01003	Streptococcus pneumon	4.57e-02	
12	139	7.9	229 4 R21613	Sequence of Plasmodium	1.14e-01	
13	139	7.9	667 31 W48760	BOP1 protein.	1.14e-01	
14	138	7.8	129 4 P40377	Sequence encoded by t	1.32e-01	
15	135	7.7	1785 23 W24790	P. faiciparum liver s	2.08e-01	
16	135	7.7	3135 11 R57474	P. faiciparum transmi	2.08e-01	

17	133	7.6	205	3	P60582	P. falciparum peptide	2.80e-01
18	130	7.4	680	3	R12540	Rabbit sperm-specific	4.39e-01
19	128	7.3	578	38	W82635	Ehrlichia sp. extende	5.91e-01
20	128	7.3	578	39	W89273	Granulocytic Ehrlich	5.91e-01
21	129	7.3	844	3	P60570	Sequence of the falc	5.09e-01
22	127	7.2	594	6	R34936	CENP-B.	6.86e-01
23	126	7.2	905	26	W31186	Human p160 polypeptid	7.95e-01
24	126	7.2	1135	26	W31185	Human p160 polypeptid	7.95e-01
25	123	7.0	754	30	W33811	Tat stimulatory facto	1.24e+00
26	121	6.9	151	25	W19101	Trypanosoma cruzi ant	1.66e+00
27	119	6.8	425	36	W80621	S. pneumoniae protein	2.22e+00
28	119	6.8	516	4	P40135	Sequence of protein A	2.22e+00
29	119	6.8	565	33	W61247	Streptococcus pneumon	2.22e+00
30	119	6.8	764	19	W01897	Nonsense-mediated MRN	2.22e+00
31	120	6.8	903	28	W37389	Human additional sex	1.92e+00
32	120	6.8	903	28	W37391	Human additional sex	1.92e+00
33	119	6.8	1089	19	W01896	Nonsense-mediated MRN	2.22e+00
34	119	6.8	1664	34	W43106	C. thermocellum OlpB	2.22e+00
35	118	6.7	205	4	R21614	Sequence of Plasmodiu	2.57e+00
36	118	6.7	664	19	W04329	Human HP-8 antigen as	2.57e+00
37	118	6.7	1706	16	R92103	Rat RIZ.	2.57e+00
38	116	6.6	258	33	W55107	Streptococcus pneumon	3.44e+00
39	114	6.5	466	38	W82398	Mouse G3BP protein.	4.59e+00
40	114	6.5	466	19	R95688	Human GAP-SH3 domain	4.59e+00
41	113	6.4	159	6	R32993	P.falciparum EB200 an	5.30e+00
42	113	6.4	159	6	R32985	P.falciparum EB200 an	5.30e+00
43	113	6.4	508	4	R23850	Vasoactive intestinal	5.30e+00
44	113	6.4	585	2	P70282	Protein A - beta-gluc	5.30e+00
45	113	6.4	831	33	W68201	M. catarrhalis strain	5.30e+00

ALIGNMENTS

RESULT 1  
ID R98747 standard; Protein; 1018 AA.  
AC R98747:  
DT 12-NOV-1996 (first entry)  
DE P. vivax ESP-1 blood stage antigen.  
KW ESP-1; blood stage antigen; diagnosis; malaria; infection;  
KW causative agent; antibody; monoclonal; polyclonal; assay.  
OS Plasmodium vivax (clone PvMB3.3.1).  
FT Key Location/Qualifiers  
FT region 129..191  
FT misc\_difference 129..137  
FT /label= repeat\_unit  
FT /note= "repeat unit characterised by the sequence  
D(L/M)PAGEE(A/T)G"  
FT domain 649..728  
FT /label= repeat\_domain  
FT /note= "contains 10 repeat units"  
FT misc\_difference  
FT /label= repeat\_unit  
FT /note= "characterised by the sequence EEVEEVP"  
PN US5532133-A.  
PD 02-JUL-1996.  
PF 02-JUN-1993; 072610.  
PR 02-JUN-1993; US-072610.  
PA (UNY ) UNIV NEW YORK STATE.  
PI Barnwell JW;  
DR WPI; 96-321110/32.  
DR N-PSDB; T34620.  
PT Antibodies to Plasmodium vivax blood stage antigens - used to  
PT diagnose malaria and to determine whether P. vivax is the species  
PT responsible for infection  
PS Example 4; Column 19-24; 22pp; English.  
CC The present sequence is that of a species-specific Plasmodium vivax  
CC malarial antigen, PvESP-1. The gene (T34620) appears to be missing a  
CC small portion of its 5' end. This protein is secreted into the plasma  
CC of a susceptible mammalian host after infection. Monoclonal/polyclonal  
CC antibodies can be utilised in assays used to diagnose malaria, as well  
CC as to determine whether P. vivax is the species responsible for the  
CC infection.

[illegible]



PS Disclosure; Fig 8B(2); 184pp; English.



PA	(STAT-) Statens Seruminst.
PI	Dziedgiew M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A, Jakobsen PH;
PR	WPI; 90-115987/15.
DR	N-PSDB; Q03875.
PT	Polypeptide(s) derived from Plasmodium falciparum antigen - used in
PT	vaccines and in production of antibodies, for diagnosis and
PT	therapy of malaria.
PS	Disclosure; Fig 8; 108pp; English.
CC	The malarial GUPR antigen is isolated from a genomic library and
CC	CC affinity purified using antibodies from malaria-immune patients. The
CC	antigen, and Abs raised therefrom may be used in diagnosis, immunisation
CC	and treatment of the disease.
CC	Sequence 783 AA;
SQ	

Query Match 9.6%; Score 169; DB 3; Length 783;  
Best Local Similarity 25.9%; Pred. No. 1.11e-03;  
Matches 50; Conservative 46; Mismatches 86; Indels 11; Gaps 10;

Db	385	eveillpeddknekveheiveveeillaedknekqgheiveveeillpeddknekqgheive	444
QY	71	EPBQVPEVEAPQNIIDEAKEIQISLLHEMVH--AEH--VEGEDLQOEDGPTGEPQOEDDE	127
Db	445	-veeillpeddknekqgheiveveeillpeddknekveheiv-eveeillpeddknekqgheiver	502
QY	128	FLMATVDDBFETLEPVSHEETEHSHVEETVSQCNDQMEEMMSEQNPSSPEVW-	186
Db	503	eellpeddknekvqheiveveeillpeddknekqgheiveveeillpeddknekqheivevee	562
QY	187	DERLHHTDD-VIYQVVEQAVY-EPLENEGI-EITEVTAP-PEDNPVEDSQ-VIVEEVS	241
Db	563	ilpeddknekvqhe	575
QY	242	IFPVEEQOEVPDP	254

RESULT	11	
ID	W61003	standard; Protein; 465 AA.
AC	W61003;	
AD		
DE	13-OCT-1998	(first entry)
DE	Streptococcus pneumoniae	encoded polypeptide.
DE	coding region;	ORF: open reading frame; antibacterial;
KW	infection; prevention;	meningitis.
OS	Streptococcus pneumoniae.	
PN	W09819889-Al.	
PN	14-MAY-1998.	
PF	27-OCT-1997;	U19226.
PR	01-NOV-1996;	US-0299330.
PA	(SMIK )	SMITHKLINE BEECHAM PLC.
PA	(SMIK )	SMITHKLINE BEECHAM PLC.
PI	Black WT, Hodgson JE,	Knowles DJC, Lonetto MA, Nicholas RO,
PI	Reid RH, Zarfos PN;	
NP	WPI: 98-286586/25.	
NP	N-PSDB: V37391.	
PT	New isolated nucleic acids from Streptococcus pneumoniae	- useful,
PT	e.g. for identifying anti-bacterial(s) for treatment and prevention	
PT	of meningitis	

Claim 11: Page 98: 130pp: English.  
The sequence is that of the polypeptide encoded by a region isolated from *S. pneumoniae*. The protein, or agonists of it, may be useful as an antibacterial for treatment or prevention of infection, specifically caused by *S. pneumoniae* (particularly meningitis) but possibly also *Helicobacter pylori* (ulcers and gastric cancer). It may be of particular use before insertion of an in-dwelling device or any other invasive procedure. The protein, or nucleic acid encoding it, can also be used in vaccines to induce a cellular and/or humoral immune response, or to screen for other antibacterials. The DNA may also contain flanking sequences that are potential sources of control elements for bacterial gene expression. Detecting a sequence encoding the protein can be used diagnostically, e.g. to detect a mutation for serotyping or classifying infectious agents.  
Sequence 465 AA;

SQ Sequence 465 AA;

Query Match	8.2%	Score 145;	DB 33;	Length 465;
Best Local Similarity	25.84;	Pred. No. 4.57e-02;		
Matches	40;	Conservative	Mismatches 69;	Indels 6; Gaps 6;
Db	95	pkveeg-kedsaapsveevgvespkeekvavkpqspdkpaeskvpepvcaakvp	153	
Qy	45	AKVLGLKERTSPVAPPPEAPHTPEEQVPVEAPQNIEDAKI-QIQSL-LHEMVH	102	
Db	154	eqvqptqaeqpsptkessgnpknedrgaeetpkgedeqaeaqekvaeepveskeetv	213	
Qy	103	AEHVEGEDLQOEDGPTCEPQOEDDEFLMATDVDDRFTEPLEPVSHE-ETESHYHV-EETV	160	
Db	214	nqvpeqkvtpavekqtetpeepkvevtslpqtt	248	
Qy	161	SODCNODMEENMSROENPDSSE-PVVDERTLHDT	194	

RESULT 12

ID R21613 standard; Protein; 229 AA.

AC R21613;

DT 23-JUL-1992 (first entry)

DE Sequence of Plasmodium falciparum immunogenic peptide IIb.

KW Immunogen; vaccine; malaria; epitope; probe; antibody.

OS Plasmodium falciparum.

PN W09203552-A.

PD 05-MAR-1992.

PR 14-AUG-1991; F00667.

PR 14-AUG-1990; FR-010363.

PA (INSP.) INST PASTEUR.

PI Muller Hill B, Kun J, Schreiber M, Gysin J, Pereira Da Silva L, Breton C;

DR WPI; 92-096896/12.

DR N-PSDB: Q21768.

PT New polypeptide(s) inducing protective antibodies - having .

PT mediator function on infected red corpuscles and useful as

PT diagnostic agents and vaccines

PS Claim 4; Fig 3: 66pp; French.

CC The polypeptides of the invention contain at least one sequence

CC having at least one epitope characteristic of a protein present on

CC the surface of red blood cells (RBC) infected with Plasmodium

CC falciparum (P.f.). They are useful as immunogens in protective

CC vaccines against malaria, and for in vitro immuno-detection of P.f.

CC in tissues or biological fluids. Nucleic acid sequences encoding

CC the polypeptides are also claimed, and so are nucleotide probes

CC contg. all or part of the nucleic acid sequences.

SO Sequence 229 AA.

```

Query Match      7.9%  Score 139;  DB 4;  Length 229;
Best Local Similarity 22.2%  Pred. No. 1.14e-01;
Matches 36;  Conservative 57;  Mismatches 57;  Indels 12;  Gaps 11;

Db      3  svseieeegsvveeieeegsvveeieeegsvveev-detelvnldelvegaptfeev 61
      ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      84  NIEDEAEQIQLLHEWVHAHVEGEDLOQEDGPTGEPOQEDDEFLLMATDVDD-RF-ETL 141
      ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db      62  eegqsvndeliedasvaeav-ee-sesitesvsgeetekg-fwie-kve-etgavtee 115
      ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      142  EPEVS-HEETEHSHVETVSDCNQDNEMEMGEQENPDSSEPPVEDERLHHTDDVTYQ 200
      ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db      116  ivqdgllteeleesvngelinkesdaeilletfeeteev 157
      ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qv      201  VYEEQAVPEP-LEN-EGEITTEVTAPPEDNPVEDSVOIVAEV 240
      ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT	13
ID	W48760 standard; Protein; 667 AA.
AC	W48760;
DT	18-AUG-1998 (first entry)
DE	BOP1 protein.
KW	Corticotropin; pituitary; BOP1; tumour suppressor; constitutive;
KW	inducible; Alzheimer's disease; nuclear transcription factor; apoptosis;
KW	cell cycle; neuronal disorder.





\*\*\*\*\*

WORLD (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:28:09 1999; MasPar time 4.58 Seconds  
Tabular output not generated. 651.905 Million cell updates/sec

Title: >US-09-040-485-2  
Description: (1-255) from US09040485.pep  
Perfect Score: 1758  
Sequence: 1 MVIALGVWTSVAVVWFELV.....IVESVIFPVVEQEQVPPDT 255

Scoring table: PAM 150  
Gap 11

Searched: 119857 seqs, 11713122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 29.732; Variance 160.216; scale 0.186

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	208	11.8	1	US-08-072-Sequence 2, Applicatio	6.34e-07
2	208	11.8	2	US-08-719-Sequence 2, Applicatio	6.34e-07
3	192	10.9	2	US-08-728-Sequence 2, Applicatio	8.69e-06
4	177	10.1	4	5231168-2 Patent No. 5231168.	9.85e-05
5	139	7.9	667	US-08-718-Sequence 2, Applicatio	3.98e-02
6	137	7.8	3052	US-08-557-Sequence 26, Applicatio	5.44e-02
7	135	7.7	3135	US-08-323-Sequence 2, Applicatio	7.36e-02
8	121	6.9	151	US-08-557-Sequence 32, Applicatio	6.18e-01
9	119	6.8	425	US-08-986-Sequence 2, Applicatio	8.33e-01
10	119	6.8	764	PCT-US95-1 Sequence 4, Applicatio	8.33e-01
11	119	6.8	764	US-08-375-Sequence 2, Applicatio	7.17e-01
12	120	6.8	903	US-08-853-Sequence 2, Applicatio	8.33e-01
13	119	6.8	1089	PCT-US95-1 Sequence 2, Applicatio	8.33e-01
14	119	6.8	1089	US-08-375-Sequence 2, Applicatio	8.33e-01
15	118	6.7	564	US-08-421-Sequence 6, Applicatio	9.67e-01
16	118	6.7	1706	US-08-399-Sequence 2, Applicatio	9.67e-01
17	118	6.7	1706	US-08-459-Sequence 2, Applicatio	9.67e-01
18	114	6.5	466	US-08-836-Sequence 5, Applicatio	1.75e+00
19	115	6.5	1702	PCT-US95-1 Sequence 9, Applicatio	1.51e+00
20	113	6.4	486	US-08-942-Sequence 3, Applicatio	2.03e+00
21	111	6.3	531	US-08-933-Sequence 9, Applicatio	2.73e+00
22	109	6.2	546	US-08-669-Sequence 3, Applicatio	3.66e+00
23	109	6.2	900	US-08-630-Sequence 62, Applicatio	3.66e+00

24	109	6.2	900	2	US-09-005-Sequence 62, Applicati	3.66e+00
25	107	6.1	251	2	US-08-766-Sequence 3, Applicatio	4.90e+00
26	107	6.1	789	1	US-08-431-Sequence 20, Applicati	4.90e+00
27	107	6.1	789	1	US-08-398-Sequence 6, Applicatio	4.90e+00
28	107	6.1	1848	3	PCT-US95-1 Sequence 4, Applicatio	4.90e+00
29	107	6.1	2101	3	PCT-US93-0 Sequence 4, Applicatio	4.90e+00
30	107	6.1	2101	1	US-08-195-Sequence 4, Applicatio	4.90e+00
31	107	6.1	2101	1	US-08-470-Sequence 4, Applicatio	4.90e+00
32	107	6.1	2101	2	US-08-467-Sequence 4, Applicatio	4.90e+00
33	107	6.1	2101	2	US-08-483-Sequence 4, Applicatio	4.90e+00
34	107	6.1	2101	1	US-08-466-Sequence 4, Applicatio	4.90e+00
35	106	6.0	182	2	US-08-466-Sequence 5, Applicatio	5.67e+00
36	106	6.0	182	3	PCT-US95-1 Sequence 5, Applicatio	5.67e+00
37	106	6.0	182	1	US-08-466-Sequence 5, Applicatio	5.67e+00
38	106	6.0	375	2	US-08-363-Sequence 7, Applicatio	5.67e+00
39	106	6.0	528	2	US-08-363-Sequence 14, Applicati	5.67e+00
40	106	6.0	530	2	US-08-363-Sequence 3, Applicatio	5.67e+00
41	106	6.0	530	2	US-08-363-Sequence 2, Applicatio	5.67e+00
42	106	6.0	530	2	US-08-363-Sequence 10, Applicati	5.67e+00
43	106	6.0	530	2	US-08-363-Sequence 8, Applicatio	5.67e+00
44	106	6.0	530	2	US-08-363-Sequence 9, Applicatio	5.67e+00
45	106	6.0	2414	3	PCT-US95-0 Sequence 2, Applicatio	5.67e+00

## ALIGNMENTS

RESULT 1  
ID US-08-072-610-2 STANDARD; PRT; 1018 AA.  
XX  
AC  
XX  
XX  
DT  
XX  
DE  
XX

Sequence 2, Application US/08072610

Sequence 2, Application US/08072610  
Patent No. 5532133

GENERAL INFORMATION:

APPLICANT: Barnwell, John

TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,

TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby and Darby

STREET: 805 Third Ave.

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022-7513

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/072,610

FILING DATE: 19930602

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Gogoris, Adda

REGISTRATION NUMBER: 29,714

REFERENCE/DOCKET NUMBER: 5986/07686

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)527-7700

TELEFAX: (212)753-6237

TELEX: 236687

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1018 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

```
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: C-terminal
CC ORIGINAL SOURCE: Plasmodium vivax
CC IMMEDIATE SOURCE:
CC CLONE: PwMB3.3.1
SQ SEQUENCE 1018 AA; 113447 MW; 5293053 CN;

Query Match 11.8%; Score 208; DB 1; Length 1018;
Best Local Similarity 32.2%; Pred. No. 6.34e-07;
Matches 69; Conservative 49; Mismatches 76; Indels 20; Gaps 18;

Db 533 LKDPAGEAVTVPSKEAPVQVPVAVGPAQEVPT-ELMQLQEDDFE-LEGTAAPAEPEGEL 590
QY 51 LKERSTSEP-AVPPEEA--E-PHT-EPEEQVPAEAPQNIIDEAKAQIQLLHEMVHAEH 105
Db 591 VLEGEGETPEEPREGTEGEVPEELEEATPEDD-FE-LE-EPTGEEVEETVEGEET-A 646
QY 106 V-EGEDLQOEDGPT-GEQQED--DEFLMATDVEDRFTLEPEVSHETSHYHVEETVS 161
Db 647 -E-GEEVEEPAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 704
QY 162 QDCNQDMEEMSEQENPDSPVVED-ERLHDTDDVTYQVVEEQAVYPLENEGIEITE 220
Db 705 VPEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 737
QY 221 VTAPPEDNPVEDSQVI-VEEVSIFPVEEQEVPP 253

RESULT 2
ID US-08-719-822B-2 STANDARD; PRT; 1018 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 2, Application US/08719822B
Sequence 2, Application US/08719822B
Patent No. 5874527
GENERAL INFORMATION: Plasmodium vivax Blood Stage Antigens
APPLICANT: Barnwell, John
TITLE OF INVENTION:
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby and Darby
STREET: 805 Third Ave.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-7513
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/719,822B
FILING DATE: 09/30/96
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 5986/17686US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)527-7700
TELEFAX: (212)753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: amino acid
TOPOLOGY: linear

CC ANTI-SENSE: NO
CC FRAGMENT TYPE: C-terminal
CC ORIGINAL SOURCE: Plasmodium vivax
CC IMMEDIATE SOURCE:
CC CLONE: PwMB3.3.1
SQ SEQUENCE 1018 AA; 113447 MW; 5293053 CN;

Query Match 11.8%; Score 208; DB 2; Length 1018;
Best Local Similarity 32.2%; Pred. No. 6.34e-07;
Matches 69; Conservative 49; Mismatches 76; Indels 20; Gaps 18;

Db 533 LKDPAGEAVTVPSKEAPVQVPVAVGPAQEVPT-ELMQLQEDDFE-LEGTAAPAEPEGEL 590
QY 51 LKERSTSEP-AVPPEEA--E-PHT-EPEEQVPAEAPQNIIDEAKAQIQLLHEMVHAEH 105
Db 591 VLEGEGETPEEPREGTEGEVPEELEEATPEDD-FE-LE-EPTGEEVEETVEGEET-A 646
QY 106 V-EGEDLQOEDGPT-GEQQED--DEFLMATDVEDRFTLEPEVSHETSHYHVEETVS 161
Db 647 -E-GEEVEEPAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 704
QY 162 QDCNQDMEEMSEQENPDSPVVED-ERLHDTDDVTYQVVEEQAVYPLENEGIEITE 220
Db 705 VPEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 737
QY 221 VTAPPEDNPVEDSQVI-VEEVSIFPVEEQEVPP 253

RESULT 3
ID US-08-728-323A-2 STANDARD; PRT; 1162 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 2, Application US/08728323A
Sequence 2, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/NSC/SKS
TELECOMMUNICATION INFORMATION:
```









QY	148	EETEHSY-HVEE-TVSQDCNQDMEEEMSEQENPDSPVPV	185
RESULT	12		
ID	US-08-853-310-2	STANDARD;	PRT; 903 AA.
XX	AC	AC	XXXXXX
XX	AC	AC	
XX	AC	AC	
XX	AC	AC	
DE	Sequence 2, Application US/08853310		
XX	Sequence 2, Application US/08853310		
CC	Patent No. 5948640		
CC	GENERAL INFORMATION:		
CC	APPLICANT: Randazzo, Filippo		
CC	TITLE OF INVENTION: Mammalian Additional Sex Combs (Asx) Acts as		
CC	NUMBER OF SEQUENCES: 4		
CC	CORRESPONDENCE ADDRESS:		
CC	ADDRESSEE: Chiron Corporation		
CC	STREET: 4560 Horton Street		
CC	CITY: Emeryville		
CC	STATE: California		
CC	COUNTRY: U.S.A.		
CC	ZIP: 94608		
CC	COMPUTER READABLE FORM:		
CC	MEDIUM TYPE: Floppy disk		
CC	COMPUTER: IBM PC compatible		
CC	OPERATING SYSTEM: PC-DOS/MS-DOS		
CC	SOFTWARE: PatentIn Release #1.0, Version #1.30		
CC	CURRENT APPLICATION DATA:		
CC	APPLICATION NUMBER: US/08/853,310		
CC	FILING DATE:		
CC	CLASSIFICATION: 514		
CC	ATTORNEY/AGENT INFORMATION:		
CC	NAME: Guth, Joseph H.		
CC	REGISTRATION NUMBER: 31,261		
CC	REFERENCE/DOCKET NUMBER: 1228.003		
CC	TELECOMMUNICATION INFORMATION:		
CC	TELEPHONE: (510) 923-3888		
CC	TELEFAX: (510) 655-3542		
CC	INFORMATION FOR SEQ ID NO: 2:		
CC	SEQUENCE CHARACTERISTICS:		
CC	LENGTH: 903 amino acids		
CC	TYPE: amino acid		
CC	STRANDEDNESS: single		
CC	TOPOLOGY: linear		
CC	MOLECULE TYPE: protein		
CC	SEQUENCE 903 AA; 97561 MW; 3803092 CN;		
QY	Query Match	6.8%; Score 120; DB 2; Length 903;	
QY	Best Local Similarity	15.6%; Pred. No. 7.17e-01;	
QY	Matches	21; Conservative 54; Mismatches 57; Indels 3; Gaps	
Db	182	LTINOMPHASQPOQNAQSQRCILLVDSNGQIIGNFLQQQRQOQQOQLLQQTLLQ	241
QY	49	LGUKERTSPVPPEPAEPHTPEQVVPVEAPQNIED-EAEQIQSLLLHEMVAEHVE	107
Db	242	AAAAQQQQQQQQHQOQQQQQQQQQSSNSLTKLT-EVALRNGTQQ-FLSPNLIAQQHQOQ	299
QY	108	GEDLQEDGGTGPQEQEDDEFLMATDVDDRRFLETEFVSHSETEHSYHVEETVSQDCNQD	167
Db	300	QQOOLEHQHQOQATAQ	314
QY	168	MEEMMSQENPDSSSE	182
RESULT	13		
ID	PCT-US95-16930-2	STANDARD;	PRT; 1089 AA.
XX	AC	AC	XXXXXX
XX	AC	AC	
XX	AC	AC	

[illegible]



\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:23:44 1999; MasPar time 12.61 Seconds  
Tabular output not generated.  
Title: >US-09-040-485-2  
Description: (1-255) from US09040485.pap  
Perfect Score: 1758  
Sequence: 1 MVIALLGWTSVAVVWFDLV.....IVEEVSIFPVEEQEVPPDT 255  
Scoring table: PAM 150  
Gap 11  
Searched: 122810 seqs, 40068593 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir60  
1:pir1-2:pir2 3:pir3 4:pir4  
Statistics: Mean 46.304; Variance 132.862; scale 0.349  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	1757	99.9	757	2	I38423	aspartyl beta-hydroxy	1.19e-231
2	744	42.3	754	1	BABOH	peptide-aspartate bet	5.82e-84
3	197	11.2	1110	2	I51116	NF-180 - sea lamprey	7.95e-10
4	192	10.9	1094	2	S49313	protein kinase - slim	3.10e-09
5	186	10.6	1073	2	A25326	ring-infected erythro	1.56e-08
6	184	10.5	304	2	S21342	ring-infected erythro	2.68e-08
7	177	10.1	411	2	S47436	flagellar antigen - T	1.73e-07
8	177	10.1	1271	2	A45555	glutamate rich protei	1.73e-07
9	165	9.4	407	1	EDBEQ3	immediate-early prote	4.00e-06
10	163	9.3	1948	2	S00485	gene 11-1 protein pre	6.70e-06
11	158	9.0	506	2	S47439	12 protein - Trypanos	2.41e-05
12	155	8.8	837	2	JN0292	antigen 332 - Plasmod	5.17e-05
13	152	8.6	302	2	S11598	ribosomal protein s3	1.10e-04
14	151	8.6	347	2	I46592	involucrin - pig	1.42e-04
15	149	8.5	589	2	S74668	hypotheical protein	2.34e-04
16	149	8.5	721	2	S29795	hypotheical protein	2.34e-04
17	149	8.5	4910	2	S64942	probable membrane pro	2.34e-04
18	147	8.4	497	1	CEECFY	cell division protein	3.84e-04
19	148	8.4	700	2	A54641	interspersed repeat a	3.00e-04
20	144	8.2	501	2	C71948	hypotheical protein	8.07e-04
21	143	8.1	772	2	I50463	protein kinase - chic	1.03e-03
22	140	8.0	505	2	B64560	poly E-rich protein -	2.15e-03
23	140	8.0	1616	2	G64242	cytadherence-accessor	2.15e-03

24	139	7.9	1018	2	S73720	cytadherence accessor	2.74e-03
25	138	7.8	607	2	S27776	80K protein (allele C	3.49e-03
26	136	7.7	304	2	S44897	ZK1236.2 protein - Ca	5.65e-03
27	135	7.7	304	1	R3HS3S	ribosomal protein S3	7.19e-03
28	136	7.7	787	2	A48819	nuclear autoantigenic	5.65e-03
29	135	7.7	3135	2	A48584	transmission blocking	7.19e-03
30	133	7.6	432	2	G71621	protein of the MAK16	1.16e-02
31	133	7.6	1558	2	B71603	RESA-H3 antigen PF809	1.16e-02
32	132	7.5	671	2	S53407	CHS5 protein - yeast	1.47e-02
33	132	7.5	1840	2	D71612	hypotheical protein	1.47e-02
34	130	7.4	680	2	A43800	nuclear autoantigenic	2.36e-02
35	130	7.4	919	1	PXZPIP	H+-transporting ATPas	2.36e-02
36	128	7.3	387	1	A43704	involucrin - western	3.77e-02
37	128	7.3	678	2	A54514	glutamic acid-rich pr	3.77e-02
38	129	7.3	834	2	D57282	ankyrin-related prote	2.98e-02
39	129	7.3	924	2	S06117	myosin heavy chain, n	2.98e-02
40	126	7.2	2007	1	B43402	myosin heavy chain-B,	2.98e-02
41	126	7.2	286	2	A34599	DNA-binding protein M	6.01e-02
42	127	7.2	572	2	S21325	probable exo-glucanas	4.76e-02
43	127	7.2	802	2	S48529	NAB3 protein - yeasts	4.76e-02
44	127	7.2	920	1	PXRZP	H+-transporting ATPas	4.76e-02
45	126	7.2	1300	2	T03166	probable immediate ea	6.01e-02

ALIGNMENTS

RESULT 1  
ENTRY I38423 #type complete  
TITLE aspartyl beta-hydroxylase - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 10-Jul-1998  
ACCESSIONS I38423  
REFERENCE I38423  
#authors Koriolth, F.; Gieffers, C.; Frey, J.  
#journal Gene (1994) 150:395-399  
#title Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase.  
#cross-references MUID:95121937  
#accession I38423  
##status preliminary; translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-757 #label RES  
##cross-references EMBL:003109; NID:g458031; PID:g458032  
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology  
FEATURE 54-75 #domain transmembrane #status predicted #label TRM  
SUMMARY #length 757 #molecular-weight 85498 #checksum 2143

Query Match	99.9%	Score 1757;	DB 2;	Length 757;
Best Local Similarity	99.6%	Pred. No. 1.19e-231;	Mismatches 1;	Indels 0; Gaps 0;
Matches	254;	Conservative		
Db	59	MVTALLGWTSVAVVWFDLV	EVVGLKGIYDADGDFD	VDVDAKVLGLKERSTSEPA 118
QY	1	MVTALLGWTSVAVVWFDLV	EVVGLKGIYDADGDFD	VDVDAKVLGLKERSTSEPA 60
Db	119	VPPEAEPHTEPEPOVVEAEPQ	IEDEAKEQIQSLHEMVHAEH	VEGEDLQQEDGPTGE 178
QY	61	VPPEAEPHTEPEPOVVEAEPQ	IEDEAKEQIQSLHEMVHAEH	VEGEDLQQEDGPTGE 120
Db	179	PQEDDEFLMATDVRFRFETLE	PEVSHSEETSHYHVEETVSQ	CNQNQNMESQENPDS 238
QY	121	PQEDDEFLMATDVRFRFETLE	PEVSHSEETSHYHVEETVSQ	CNQNQNMESQENPDS 180
Db	239	SEPVVEDRLHDDDDVYQVYE	QAVYEPLENEGIEITEVTAP	EDNPVEDSQVIVEEV 298
QY	181	SEPVVEDRLHDDDDVYQVYE	QAVYEPLENEGIEITEVTAP	EDNPVEDSQVIVEEV 240
Db	299	SIFPVEEQEVPPET	313	
QY	241	SIFPVEEQEVPPDT	255	

```

RESULT 2
ENTRY BABOH #type complete
TITLE peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine
ALTERNATE_NAMES aspartyl (asparaginyl) beta-hydroxylase
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 31-Dec-1993 #sequence_revision 10-Feb-1995 #text_change
29-May-1998

ACCESSIONS A42969; A39470; B39470; C39470; S27948
REFERENCE A42969
#authors Jia, S.; VanDusen, W.J.; Diehl, R.E.; Kohl, N.E.; Dixon,
R.A.; Elliston, K.O.; Stern, A.M.; Friedman, P.A.
#journal J. Biol. Chem. (1992) 267:14322-14327
#title CDNA cloning and expression of bovine aspartyl (asparaginyl)
beta-hydroxylase.
#cross-references MUID:92332546
#accession A42969
#molecule_type mRNA
#residues 1-754 #label JTA
#cross-references EMBL:M91213; NID:g162693; PID:g162694
#experimental_source brain
#note sequence extracted from NCBI backbone (NCBIP:108534)
REFERENCE A39470
#authors Wang, Q.; VanDusen, W.J.; Petroski, C.J.; Garsky, V.M.;
Stern, A.M.; Friedman, P.A.
#journal J. Biol. Chem. (1991) 266:14004-14010
#title Bovine liver aspartyl beta-hydroxylase. Purification and
characterization.
#cross-references MUID:91310689
#accession A39470
#molecule_type protein
#residues 289-328 #label WAN
#accession B39470
#molecule_type protein
#residues 615,'X',617-630,'XX',633-634,'X',636,'XX',639-641
#label WA2
#accession C39470
#molecule_type protein
#residues 311-347,'X',349,'X',351-373,'X',375-379,'X',381-382
#label WA3
COMMENT This enzyme uses ferrous iron as a cofactor, and while
beta-hydroxylating the peptidyl-aspartate substrate converts
alpha-ketoglutarate to succinate and releases carbon dioxide.
COMMENT Aspartic acid and asparagine residues in the EGF homology domain of
certain plasma proteins serve as the peptidyl-aspartate
substrate.
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase;
tetrairicopeptide repeat homology
KEYWORDS glycoprotein; oxidoreductase; transmembrane protein
FEATURE
2-56 #domain intracellular #status predicted #label INC\
57-78 #domain transmembrane #status predicted #label TRM\
289-754 #product peptide-aspartate beta-dioxygenase, 56k form
#status predicted #label 56K\
311-754 #product peptide-aspartate beta-dioxygenase, 52k form
#status predicted #label 52K\
337-370 #domain tetrairicopeptide repeat homology #label TT1\
371-404 #domain tetrairicopeptide repeat homology #label TT2\
13,96,466,702 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 754 #molecular-weight 84998 #checksum 9667
Query Match 42.3%; Score 744; DB 1; Length 754;
Best Local Similarity 56.2%; Pred. No. 5.82e-84;
Matches 140; Conservative 33; Mismatches 52; Indels 24; Gaps 8;

Db 62 MVIALLGWTSVAVVWFDLYEEVLAKRPNLSEVLQKGIYADGGDFDVA 121
|||||
Qy 1 MVIALLGWTSVAVVWFDLYEYL-----GKLGIDYADGGDFDVA 45
|||||

Db 122 KVLGLKEKPAKPVTPPEADMYPLNEDVLESGRONIEDVEVQVSL-DETVYSEP 180
|||||

```

---

```

Qy 46 KVLGLKERSTSEPAVPPPEAEPTPEEQVPVEAPQNIIDEAKQIQSLLEHMYHAEH 105
Db 181 --GENLPQEPGPAELQDDHVF-VGSADDDRPYPMGTGAVHEETEDSYHIEETASPAY 237
|||||
Qy 106 VEGEDLQGE-DGTEGPEQOEDDEFLMATDVRFFETLEPEVSHETSHYHVEETVSQDC 164
|||||
Db 238 SODMDDMYEQENPSSEPVVVDAAERTYQETDDVTYRDYDQD--HAVDNSNTILEEPH 295
|||||
Qy 165 NODMEEMSEQENPSSEPVV-ED-ERLHDDTDDVTYQVEQAVYEPLENEGIEITEVT 222
|||||
Db 296 MPPEAEQOE 304
|||
Qy 223 APPEDNPVE 231
|||

RESULT 3
ENTRY 151116 #type complete
TITLE NF-180 - sea lamprey
ORGANISM #formal_name Petromyzon marinus #common_name sea lamprey
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
13-Sep-1996
ACCESSIONS 151116
REFERENCE 151116
#authors Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
#journal Brain Res. Mol. Brain Res. (1995) 29:43-52
#title The single lamprey neurofilament subunit (NF-180) lacks
multiphosphorylation repeats and is expressed selectively
in projection neurons.
#cross-references MUID:95287814
#accession 151116
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1110 #label JAC
#cross-references EMBL:U19361; NID:g632548; PID:g632549
SUMMARY #length 1110 #molecular-weight 123817 #checksum 5627

Query Match 11.2%; Score 197; DB 2; Length 1110;
Best Local Similarity 25.8%; Pred. No. 7.95e-10;
Matches 51; Conservative 50; Mismatches 91; Indels 6; Gaps 6;

Db 453 SPISAQLDLDLELAQEEV-MEAKAAPVVSAAEKDEEEEEEEEEEEEEEEEDR 511
|||||
Qy 57 SEPAVPPPEAEPTPEEQVPVEAPQNIIDEAKQIQSLLEHMYHAEHVEGEDLQEDG 116
|||||
Db 512 GRKEGEAEAE-EAEEVEKE-EAAEAEVEAEAEET-EAAAEAEAEAEAEAEAEAE 568
|||||
Qy 117 PRGEQOEDDEFLMATDVRFFETLEPEVSHETSHYHVEETVSQDCNQDMEMMSEQE 176
|||||
Db 569 EAEAEAEVEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 626
|||||
Qy 177 NPDSSEPVVDDRLHDDTDDVTYQVEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVI 236
|||||
Db 627 VEEGEAEAEAEAEAEAE 644
|||||
Qy 237 VEEVSIFPVEEQEVPDP 254
|||||

RESULT 4
ENTRY 549313 #type complete
TITLE protein kinase - slime mold (Dictyostelium discoideum)
ORGANISM #formal_name Dictyostelium discoideum
DATE 16-Feb-1995 #sequence_revision 12-May-1995 #text_change
19-Dec-1997
ACCESSIONS S52076; S49313
REFERENCE S52076
#authors Wetterauer, B.W.; Hamker, U.; von Haeseler, A.; MacWilliams,
H.K.; Simon, M.N.; Veron, M.
#journal Biochim. Biophys. Acta (1995) 1265:97-101
#title A protein kinase from Dictyostelium discoideum with an
unusual acidic repeat domain.
#cross-references MUID:95161460
#accession S52076
#status preliminary

```

QY		110	DLOQDGGPTGEPOEDDEFMLATVDVDRFETLPEPVSHSEETHSVHVEVTVSQCNDQMDE	169		
Db		992	ENVEENVEEYEENVEEVENVEEVENVEEVN- EENVEEVN- EEVE- ENVP- ENVEEVENVEEN	1047		
QY		170	EMKSEQNPDSPPVVE-DERLHHDDTDVTVQVYESQAAYEPLENEGIIITEITWAPPDN	228		
Db		1048	VENVEEVENVEEYDENVEEHNE	1069		
QY		229	PVEDSQIVIVEESIFPVEEQOE	250		
RESULT		6	S21342 #type fragment ENTRY ring-infected erythrocyte surface antigen - Plasmodium falciparum (fragment) TITLE #formal_name plasmodium falciparum ORGANISM 20-Feb-1995 #sequence_revision 20-feb-1995 #text_change DATE 31-Oct-1997 ACCESSIONS S21342 REFERENCE S21342 #authors Kun, J.; Kun, J. #submission submitted to the EMBL Data Library, November 1990 #description The sequence of the 3' region of RESA of the Plasmodium falciparum strain Palo Alto.			
#accession		S21342 preliminary				
#status		molecule_type DNA				
#molecules		1-304 ##label KUN				
#residues		##cross-references EMBL:X53124; NID:g9958; PID:g9959				
CLASSIFICATION		#superfamily ring-infected erythrocyte surface antigen; dnaJ amino-terminal homology				
KEYWORDS		surface antigen				
SUMMARY		#length 304 #checksum 5924				
Query Match		10.5%; Score 184;	DB 2:	Length 304;		
Best Local Similarity		27.5%;	Pred. No.	2.68e-08;		
Matches		52; Conservative	49; Mismatches	76; Indels	12; Gaps	11;
Db		127	ENVPERHYQHNAEANVERHDAAENVEHADAEENEVHDAENAEE- ENVE- ENVEEVEEVENVEEVN	184		
QY		64	EAEAPHTPEQEYPVAEPQ-NIEDEAKIQISLLHMVHAHVHEGEDAQEDGPGTEPQ	122		
Db		185	BEVVNVVNVEEVENVEEVEEV- ERNER-- NVSENVEEVENVEEVN- EENVEEVNVE	240		
QY		123	QEDDFELMATDVDDRFTLPESHETSYSHEETTSVSDCQNOMMEAMSEQENDPSSE	182		
Db		241	ENVEEVEEYDEENVE-EV-EEN-VNEENVE- ENVE- ENVEEVENVEEVENVEEVEEDE	295		
QY		183	PWEEDERLHDDTDDVTVQVYESQAAYEPLENEGIIITEITWAPPDPVEDSQIVIVEESI	242		
Db		296	ENVEEHNGI	304		
QY		243	FVVEEQEV	251		
RESULT		7	S47436 #type fragment ENTRY flagellar antigen - Trypanosoma brucei (fragment) TITLE #formal_name trypanosoma brucei ORGANISM 13-Jan-1995 #sequence_revision 30-Jan-1998 #text_change DATE 30-Jan-1998 ACCESSIONS S47436 REFERENCE S47436 #authors Imboden, M.; Mueller, N.; Hemphill, A.; Mattioli, X.Y.Z.; Seebeck, T. #submission submitted to the EMBL Data Library, August 1994 #description Repetitive proteins from the flagellar cytoskeleton of African Trypanosomes are diagnostically useful antigens.			
#accession		S47436 molecule_type mRNA				
#residues		1-411 ##label IMB				
#cross-references		EMBL:Z36281; NID:g530358; PID:g530359				
#experimental_source		strain stock TREU 1285				



RESULT	9	
ENTRY	EDBEQ3	#type complete

```

#journal      EMBO J. (1988) 7:1129-1137
#title       The 11-1 gene of Plasmodium falciparum codes for distinct
              fast evolving repeats.
#cross-references MUID:88296416
#accession   S00485
#molecule_type DNA
#residues    1-1315;1316-1485;1486-1657;1658-1729;1730-1948 ##label
              SCH
##cross-references EMBL:X07453
COMMENT      This protein is associated with the membrane of red blood cells at
              the schizont stage of infection by Plasmodium falciparum.

GENETICS
#gene        11-1
#introns     71/3
#keywords    tandem repeat
FEATURE
1-71         #domain signal sequence #status predicted #label SIG\
72-1948      #product gene 11-1 protein (fragments) #status predicted
              #label MAT
SUMMARY      #length 1948 #checksum 7088

Query Match          9.3%; Score 163; DB 2; Length 1948;
Best Local Similarity 25.5%; Pred. NO. 6.70e-06;
Matches 51; Conservative 47; Mismatches 93; Indels 9; Gaps 9;

Db      809 VIPEIVEEVLPEEVVK-EVIPLEVEEVIPEEVEEVPPEEI-VPEVPEELVEEMKPEE 866
QY      61 VPPEASPTPEEQVPVPEAFQNTDEAK-EQISLHEMVHAEHVEGEDQQEDGPT 119
        :: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      867 VFEEVISEELVEEMKPEEVEEVTPEKVVEEVPPEELVEEMKPEEVEEVLVEEVI 926
QY      120 EPQGE-DDEFLMATDVRDFTLEFVEVSEETEHYHVEETVSDCNQDM-EMMSEQEN 177
Db      927 PEEVVEEIIPEEVHAEKHEELFKLVPVEVEEIIIPKLEVPVPEEVLPEELIEVPIE 986
QY      178 PDSS-EPVVEDERLHDDDTYQ-VVEEQAVYPLENEGIIITEVATPDNPVED-SQ 234
Db      987 EIIIEV-VIPEEVEEVIPE 1005
QY      235 VIVEEVSIFFPVEEQEVPDP 254

```

[illegible]

```
D6      201 PV-T-DTEVAKEPTDSEVIPEKIIPDTAAASQPADLTKABELDPVADVEAKEPFT 338
        :|::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy     137 RFETLEPVSHETEHSYHYEETSVDQCNDMEEMKSEQENPD$-SEPVVVEDRLHHDTD 195
        !|::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     339 D-S-EVPIPEKIIPDTAAAQAPADLTAKBELDEPVADVEAKEPFTDSEVIPEKIEP 394
        !|::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy     196 DVTYQVYEQAAYEPLENEGIETEVTPAPE-DNPVEDSQIVVEVSIFPVEEQEVP 252
        :|::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT    12
ENTRY     JN0292          #type fragments
TITLE     antigen 332 - Plasmodium falciparum (fragments)
ORGANISM   formal_name Plasmodium falciparum
DATE       09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change
              22-Nov-1996

ACCESSIONS JN0292
REFERENCE   JN0292
            Mattei, D.; Scherf, A.
            Gene (1992), 110:71-79
            The Pf332 gene of Plasmodium falciparum codes for a giant
            protein that is translocated from the parasite to the
            membrane of infected erythrocytes.
#cross-references MUID:92184117
#accession      JN0292
                #status         nucleic acid sequence not shown
                #molecule_type DNA
                ##residues       1-837 #label MAT
                ##cross-references GB:M69161; GB:M69162; GB:M69163; GB:M69164
                                in the authors' translation 38-val, 458-Ser and 623-Val
                ##note           were missed
COMMENT      This protein is located in the erythrocyte cytoplasm.
GENISTICS
            Pfl32
            #gene
            #map_position 11
SUMMARY      #length 837 #checksum 9751
```

[illegible]

```

RESULT 13
ENTRY
  TITLE
  ALTERNATE_NAMES
  ORGANISM
  DATE
  ACCESSIONS
  REFERENCE
  #authors
  #journal
  #title
  #type complete
  ribosomal protein S3 - Halobacterium halobium
  ribosomal protein HS4
  #formal_name Halobacterium halobium
  31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
  18-Jun-1993
  S11598
  S11598
  S11597
  Spiridonova, V.A.; Akhmanova, A.S.; Kagramanova, V.K.;
  Koepke, A.K.E.; Mankin, A.S.
  Can. J. Microbiol. (1989) 35:153-159
  Ribosomal protein gene cluster of Halobacterium halobium
  nucleotide sequence of the genes coding for S3 and L22
  equivalent ribosomal proteins.

```

RESULT	15.
ENTRY	S74668
TITLE	#type complete hypothetical protein sll1665 - <i>Synechocystis</i> sp. (strain PCC 6803)
ORGANISM	#formal_name <i>Synechocystis</i> sp.

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Result No.	Query Match	Score	Length	ID	Description	Pred. No.	
1	99.9	1757	757	1	ASPH_HUMAN	ASPARTYL/ASPARAGINYL B	9.54e-25
2	42.3	744	734	1	ASPH_BOVIN	ASPARTYL/ASPARAGINYL B	1.31e-92
3	186	10.6	1073	1	RESA_PLAFF	RING-INFECTED ERYTHROC	1.38e-09
4	184	10.5	304	1	RESA_PLAFF	RING-INFECTED ERYTHROC	2.47e-09
5	165	9.4	407	1	IE68_HSVSA	IMMEDIATE-EARLY PROTEI	5.81e-07
6	152	8.6	302	1	RS3_HALHA	30S RIBOSOMAL PROTEIN	2.15e-05
7	151	8.6	347	1	INVO_PLIG	INVOLUCRIN.	2.82e-05
8	149	8.5	721	1	YCF2_OENPI	HYPOTHETICAL PROTEIN (	4.86e-05
9	147	8.4	497	1	FTSY_ECOLI	CELL DIVISION PROTEIN	8.35e-05
10	148	8.4	615	1	DNAK_THETH	DNAAK PROTEIN (HEAT SHO	6.37e-05
11	140	8.0	1616	1	P200_MYCGE	PROTEIN P200.	5.41e-04
12	139	7.9	1018	1	HMW1_MYCPN	CYTADHERENCE HIGH MOLE	7.04e-04
13	136	7.7	304	1	CEC1_CAEEL	CEC-1 PROTEIN.	1.54e-03
14	135	7.7	304	1	RS3_HALWA	30S RIBOSOMAL PROTEIN	2.00e-03
15	135	7.7	787	1	NASP_HUMAN	NUCLEAR AUTOANTIGENIC	1.54e-03
16	135	7.7	3135	1	S230_PLAFO	TRANSMISSION-BLOCKING	2.00e-03
17	132	7.5	671	1	CHS5_YEAST	CHITIN BIOSYNTHESIS PR	4.35e-03
18	132	7.5	1941	1	YRM8_CAEEL	HYPOTHETICAL 216.3 KD	4.35e-03
19	130	7.4	680	1	NASP_RABIT	NUCLEAR AUTOANTIGENIC	7.25e-03
20	130	7.4	919	1	PHAL_SCHPO	PLASMA MEMBRANE ATPASE	7.25e-03
21	128	7.3	387	1	INVO_TARBA	INVOLUCRIN.	1.21e-02
22	128	7.3	678	1	GARP_PLAFF	GLUTAMIC ACID-RICH PRO	1.21e-02
23	128	7.3	1976	1	MYSO_HUMAN	MYOSIN HEAVY CHAIN NO	1.21e-02



```

KUN J., LEET M., ANTHONY R.L., KUN J.E., ANDERS R.F.:
"Plasmodium falciparum: a region of polymorphism in the 3' end of the
gene for the ring-infected erythrocyte surface antigen.";
EXP. PARASITOL. 78:418-421(1994).
CC -1- FUNCTION: RESA MAY DISRUPT THE NORMAL INTERMOLECULAR INTERACTIONS
OF THE CYTOPLASMIC DOMAIN OF BAND 3 AND THEREBY FACILITATE THE
INVAGINATION OF THE RED CELL MEMBRANE WHICH IS NECESSARY FOR THE
FORMATION OF THE PARASITOPHOUS VACUOLE.
CC -1- SUBCELLULAR LOCATION: PROBABLY LOCATED ON THE CYTOPLASMIC FACE OF
THE MEMBRANE WHERE IT ASSOCIATES WITH COMPONENTS OF THE MEMBRANE
SKELETON.
CC -1- SIMILARITY: THE N-TERMINAL SEQUENCE OF BAND 3 SHOW HOMOLOGY WITH
THE REPEAT SEQUENCES OF RESA.
CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outpost -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X55124; G9959; ..
DR PROSITE; PS00636; DNAJ_1; PARTIAL.
DR PROSITE; PS50076; DNAJ_2; PARTIAL.
DR PFAM; PF00226; DnaJ; PARTIAL.
DR MALARIA; ANTIGEN; GLYCOPROTEIN; REPEAT.
FT NON_TER      1      1
FT CARBOHYD     18     18 POTENTIAL..
FT CARBOHYD     22     22 POTENTIAL..
FT NON_TER      304    304
SQ SEQUENCE      304 AA; 35225 MW; EB8E9BB4 CRC32;

Query Match          10.5%; Score 184; DB 1; Length 304;
Best Local Similarity 27.5%; Pred. No. 2.47e-09;
Matches              52; Conservative 49; Mismatches 76; Indels 12; Gaps 11.

Db   127 ENVPEHVQHNAENVHDAENVEHDAENVEHDAENAE-ENVE-ENVEEVENVEENV 184
QY   64 EEAEPTPEEQVPVAEPQ-NIEDEAKEIQISLLHEWHAHVHGEDLQEDGPTGEQ 122
Db   185 EENVEENVEENVEENVEENVEEV-EENVE--NVEENVEENVEENVEENVEENVE 240
QY   123 QEDDFELMATVDVRPFTLEPVSHETSRYHVEETVSQDCNQDMSEEMSEQENPDSE 182
Db   241 EENVEENVEEDEVNE-EV-EEN-VENVE-ENVE-ENVEENVEENVEENVEEYDE 295
QY   183 PVVEDERLHRHTDDVTQVYEQAQVPLENEGIEITEVTAPPEDNPVEDSQVIVEVSI 242
Db   296 ENVEEHNGI 304
QY   243 FPVEEQEV 251

RESULT 5
ID IE68_HSVSA STANDARD; PRT; 407 AA.
AC QO1042;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE IMMEDIATE-EARLY PROTEIN.
GN 73 OR ECLF1.
OS HERPESVIRUS SAIMIRI (STRAIN 11).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92333688.
RA ALBRECHT J.-C., NICHOLAS J., BILLER D., CAMERON K.R., BIESINGER B.,
RA NEWMAN C., WITTMANN S., CRAYTON M.A., COLEMAN H., FLECKENSTEIN B.,
RA HONESS R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
```

```

RL J. VIROL. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92230228.
RA NICHOLAS J., CAMERON K.R., COLEMAN H., NEWMAN C., HONESS R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT organization between HVS and Epstein-Barr virus."
RL VIROLOGY 188:296-310(1992).
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND
CC HSV-2 IE-68 (US1), EBV-1 65, EBV-4 (ORF4), PRV RSP40, AND VZV 63.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64346; G60394; -
DR EMBL; M86409; G331002; -
DR EMBL; S76368; G243353; -
DR PIR; G36813; EDBEQ3.
DR PIR; S20244; S20244.
DR HSP; P53041; 1A17.
KW EARLY PROTEIN.
FT DOMAIN 60 241
FT SEQUENCE 407 AA; 46617 MW; 14FCA9CF CRC32;
SQ
Query Match 9.4%; Score 165; DB 1; Length 407;
Best Local Similarity 20.7%; Pred. No. 5.81e-07;
Matches 43; Conservative 55; Mismatches 105; Indels 5; Gaps 5;
Db 40 LKPGDDDDINTHQQAALTEQRREVEEGEREGEREGEREGEREGEREGERE 99
|| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 51 LKERSTSPAPPEAEPTPEEQVPVEAPONTDEAKQIQSLHWMVHA-EHVGE 109
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 100 EAEKEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 159
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 DLQEDGPTGPQEDDFLMTDVRFTLEPEVSHVETSHVETVSDQCNDWE 169
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 160 EEAEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 218
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 EMMSEQNPDSPEPVVE-DERLHDTDVTVQVVEQAVYPLENGEITEVTAP 227
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 219 EAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 245
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 NPVEDSQVIVEEVSIFPVEEQEVPPT 255
: : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 6
ID RS3_HALHA STANDARD; PRT; 302 AA.
AC P15009; 024784;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 30S RIBOSOMAL PROTEIN S3 (HS4) (HHAS3).
OS HALOBACTERIUM HALOBIVM.
OC ARCHAEA; EURVARCHAOTA; HALOBACTERIALES; HALOBACTERIACEAE;
OC HALOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89248673.
RA SPIRIDONOVA V.A., AKHMANOVA A.S., KAGRAMANOVA V.K., KOEPKE A.K.E.,
RA MANKIN A.S.;
RT "Ribosomal protein gene cluster of Halobacterium halobium: nucleotide
RT sequence of the genes coding for S3 and L29 equivalent ribosomal
RT proteins."
RL CAN. J. MICROBIOL. 35:153-159(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97031049.

```

---

```

RA MIYOKAWA T., URAYAMA T., SHIMOOKA K., ITOH T.;
RT "Organization and nucleotide sequences of ten ribosomal protein genes
RT from the region equivalent to the S10 operon in the archaeobacterium,
RT Halobacterium halobium."
RL BIOCHEM. MOL. BIOL. INT. 39:1209-1220(1996).
RN [3]
RP SEQUENCE OF 1-16 FROM N.A.
RX MEDLINE; 89211383.
RA MANKIN A.S.;
RT "The nucleotide sequence of the genes coding for the S19 and L22
RT equivalent ribosomal proteins from Halobacterium halobium."
RL FEBS LETT. 246:13-16(1989).
CC -!- SIMILARITY: BELONGS TO THE S3P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X14967; G43552; -
DR EMBL; AB006961; D1023138; -
DR PIR; S11598; S11598.
DR PROSITE; PS00548; RIBOSOMAL_S3; 1.
DR PFAM; PF00013; KH-domain; 1.
DR PFAM; PF00189; S3_C; 1.
KW RIBOSOMAL PROTEIN.
FT CONFLICT 32
FT SEQUENCE 302 AA; 33101 MW; 8F87FDC7 CRC32;
SQ
Query Match 8.6%; Score 152; DB 1; Length 302;
Best Local Similarity 24.0%; Pred. No. 2.15e-05;
Matches 36; Conservative 42; Mismatches 61; Indels 11; Gaps 11;
Db 154 NGEPAEEI-VDHGKGVANMKGITGVNKKIIPN-AELPDDFEIQEDADIEDLVDEAEA 211
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 33 DAGDGDFDVDVDAK-V-LIGLKERSTSEPAVPEAEAPTEPEEQVPVEAEQONIED-EA 89
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 GDLBELL-EGEDAD-AEDADADAAAPSEPADPEDEVIETDDDDVEELDELADAVEG 269
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 KEQISLHWMVHAHEVEGEDLQEDGPTGEFQQ-EDDEFL-MATVDVDRFETLEPEVSH 147
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 EDEDFSDVDDEAA-DTAEIDLDEMDDED 298
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 148 EETESY-HVEETVSQDCNQDMEMMSEQE 176
: : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 7
ID INVO_PIG STANDARD; PRT; 347 AA.
AC P18175;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE INVOLUCRIN.
GN IVL.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90348475.
RA TSENG H., GREEN H.;
RT "The involucrin genes of pig and dog: comparison of their segments of
RT repeats with those of prosimians and higher primates."
RL MOL. BIOL. EVOL. 7:293-302(1990).
CC -!- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC -!- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.

```

CC This SWISS-PROT entry is copyright. It is produced through a collabor

CC This SW  
CC between

CC This SW  
CC between





```
DR EMBL; U02245; G407264; --
DR EMBL; U02245; G407265; ALT_FRAME.
DR EMBL; U02175; G406408; --
DR EMBL; U02126; G409895; --
DR TIGR; MG386; --
KW CYTADHERENCE; STRUCTURAL PROTEIN; REPEAT.
FT DOMAIN 1205 1389 2 X 32 AA REPEAT.
FT REPEAT 1205 1236 2-1.
FT REPEAT 1358 1389 2-2.
FT DOMAIN 891 1389 2 X 26 AA REPEAT.
FT REPEAT 1161 1186 2-1.
FT REPEAT 1310 1339 2-2.
FT CONFLICT 256 256 P -> S (IN REF. 2).
FT CONFLICT 304 304 S -> F (IN REF. 2).
SQ SEQUENCE 1616 AA; 185678 MW; F041E283 CRC32;

Query Match 8.0%; Score 140; DB 1; Length 1616;
Best Local Similarity 22.1%; Pred. No. 5.41e-04;
Matches 46; Conservative 54; Mismatches 99; Indels 9; Gaps 9;

Db 1156 QKVISEPQVQPGAEVFAEAKFSPVSDQSPVLEEVQVQPEIQPVESQPEA 1215
QY 53 ERSTSEPAVPPPEAEPTPEEQVPAEPQNIEDAEKEIQSLHENVHAEHVEGEDLQ 112

Db 1216 TFDVQPEQTPQAEKFDSPVETVQPEFSS-EPTQQHVSEASDFPNYDFPNYDFDQ 1274
QY 113 QEDGPTGEQEDDEFLMATD-VDDR-FETLEPEVSHEETSHVHETVSDQCNQDMEE 170

Db 1275 PSVDSO-LQSEPDYVDENYDFENYDEPNYEEKSPFQPEVQPGAEVFAEAK 1333
QY 171 MMSQENPDSPSEVVEDEHLHDDVTYQV-YE-EQAVYEP-LENE-GIEITVETAPPE 226

Db 1334 FDSPVESQDSQPELLEEVQVQPEIQP 1361
QY 227 -DNPVEDSQVIVEEVSIFPVEEQQEVPP 253

RESULT 12
ID HMW1.MYCPN STANDARD; PRT; 1018 AA.
AC Q50365; Q50348; Q50349;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTADHERENCE HIGH MOLECULAR WEIGHT PROTEIN 1 (CYTADHERENCE ACCESSORY
DE PROTEIN 1).
GN HMW1.
OS MYCOPLASMA PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-16; 176-182 AND 188-198.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE; 96257187.
RA DIRKSEN L.B., PROFT T., HILBERT H., PLAGENS H., HERRMANN R.,
RA KRAUSE D.C.
RT "Sequence analysis and characterization of the hmw gene cluster of
RT Mycoplasma pneumoniae."
RL GENE 171:19-25(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE; 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERRMANN R.
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC -1- STABILIZES THE SHAPE OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS
CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
CC MEMBRANE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; L38997; G639790; --
CC EMBL; AE000038; G1674076; --
CC EMBL; Z32661; G474076; --
CC EMBL; Z32662; G474078; --
KW CYTADHERENCE; STRUCTURAL PROTEIN.
FT CONFLICT 198 198 E -> I (IN AA SEQUENCE).
SQ SEQUENCE 1018 AA; 112214 MW; 41583DE5 CRC32;

Query Match 7.9%; Score 139; DB 1; Length 1018;
Best Local Similarity 24.8%; Pred. No. 7.04e-04;
Matches 51; Conservative 44; Mismatches 97; Indels 14; Gaps 13;

Db 270 VEPQAVQTQPIPAEQSAVLEQPEPVAEVSQEMVQPEAAAEPTVTEAQOTEPVVTETAE 329
QY 61 VPPEAEPTHE-PEQVPVPAEPQNIEDAEKEIQS-LLHEMV-HAEHVEGEDLQEDGP 117

Db 330 IT-PQVTEPVVAVVEHQPEAVAEPLPVPVAVAGVSELIPTEQVQPEVVVESTP-VAEYQ 387
QY 118 TGEFQEDDEFLMATD-DVF-ETL--EPEVS-HEETSHVHVEETVSQDCNQDMEMM 172

Db 388 SEMQVQVAVPEPTVEPQPEQVQPEVITTPVAVSLEVEQPEVQVVEQVQPEPQETP 447
QY 173 SEQENPDSS-EPVVEDEHLH-DTDDVTYQVYEQAVYE-PLENEGIEITVETAPEDNP 229

Db 448 VEVOPEVWETVQEAETQVVEPQ 473
QY 230 VE-DSQVIVEEVSIFPVEEQQEVPPD 254

RESULT 13
ID CEC1.CAEEL STANDARD; PRT; 304 AA.
AC P34618;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CEC-1 PROTEIN.
GN CEC-1 OR ZK1236.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WORLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL NATURE 368:32-38(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 'CHROMO' DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

[illegible]

Db 470 QMKEETEGBEDDKENDKT--EEMPNDVLENKSLQNEEEEEIGNLELAWMDLAKI 527  
QY 75 QVPVEAEQNIIEDEAKEQIQSLLENHMHAEHVHVEDLQQ-EDGPTGEPOQEDDEFLMATD 133  
Db 528 IFKRQETKEAQLYAAQAHKLK-GEVSVESENYVQAVEEFQS 567  
QY 134 VDDRFETLEPEVSHEETEHSYHVEETV-SQDCNQDMEEMMS 173

Search completed: Thu Oct 21 15:25:51 1999  
Job time : 40 secs.

(TM)

Result No.	Query		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	364	20.7	210	6	Q28264	JUNCTIONAL SARCOPLASMI	1.51e-28
2	197	11.2	1110	13	Q31235	NF-180.	8.24e-09
3	192	10.9	1094	5	Q23915	PROTEIN KINASE.	2.91e-08
4	192	10.9	1162	14	Q98148	ORF73 HOMOLOG.	2.91e-08
5	190	10.8	930	5	Q17339	T23E7.2B PROTEIN.	4.82e-08
6	187	10.6	1089	14	Q40947	ORF 73.	1.02e-07
7	177	10.1	411	5	Q26766	FLAGELLAR ANTIGEN (FRA	1.22e-06
8	177	10.1	1271	5	Q25860	GLUTAMATE RICH PROTEIN	1.22e-06
9	174	9.9	1262	5	Q20684	F52H3.7 PROTEIN.	2.55e-06
10	170	9.7	913	13	Q13099	MIDDLE MOLECULAR WEIGH	6.77e-06
11	158	9.0	506	5	Q26767	I2 PROTEIN (FRAGMENT).	1.21e-04
12	156	8.9	880	5	Q17338	T23E7.2B PROTEIN.	1.94e-04
13	155	8.8	345	5	P91249	SIMILAR TO COLLAGEN.	2.46e-04
14	155	8.8	837	5	Q25751	A332 ANTIGEN (FRAGMENT	2.46e-04
15	152	8.6	543	5	Q27043	POLYMORPHIC IMMUNODOMI	4.98e-04
16	151	8.6	2109	5	Q76416	H05O09.1 PROTEIN (FRAG	6.29e-04
17	149	8.5	589	2	P72805	HYPOTHETICAL 63.6 KD P	1.00e-03
18	150	8.5	718	13	Q73619	NUCLEAR PROTEIN.	7.94e-04
19	149	8.5	4910	3	Q12139	SIMILARITY NEAR N-TERM	1.00e-03
20	148	8.4	700	5	Q25894	INTERSPERSED REPEAT AN	1.26e-03



[illegible]

Db 182 EPOOV-PAEAQPEAQPE 198  
 QY 237 VEEVSIFPVEEQVEPPD 254

## RESULT 8

ID Q25860 PRELIMINARY; PRT; 1271 AA.  
 AC Q25860;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUTAMATE RICH PROTEIN.  
 GN GLURP.  
 OS PLASMODIUM FALCIPARUM.  
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92131041.  
 RA BORRE M.B., DZIEGIEL M., HOGH B., PETERSEN E., RIENECK K., RILEY E.,  
 RA METS J.F., AIKAWA M., NAKAMURA K., HARADA M., WIND A., JAKOBSEN P.H.,  
 RA COWLAND J., JEPSEN S., AXELSEN N.H., VUUST J.;  
 RT "Primary structure and localization of a conserved immunogenic  
 RT Plasmodium falciparum glutamate rich protein (GLURP) expressed in  
 RT both the preerythrocytic and erythrocytic stages of the vertebrate  
 RT life cycle."  
 RL MOL. BIOCHEM. PARASITOL. 49:119-131(1991).  
 DR EMBL; M59706; G160312;  
 SQ SEQUENCE 1271 AA; 145476 MW; 82E7B8D5 CRC32;

Query Match 10.1%; Score 177; DB 5; Length 1271;

Best Local Similarity 24.1%; Pred. No. 1.22e-06;

Matches 47; Conservative 54; Mismatches 82; Indels 12; Gaps 10;

Db 781 EKKNSQSVVEIPWNE-E-FEDVHTQOLD-LDHKTVDPEIVEEIPSELHENEVAHP 837  
 QY 64 EEAEPHTEPEEQVPEASPNQDEAKQIQSLLEHMHAEHVEGEDLQOE--DGPTGEP 121  
 Db 838 EIVELEEVFPNNQNEQNEEDDKSHIQHIEIVEEILPED-DKN-EKVEHIEIVE 895  
 QY 122 QOED-DEFLMATDVRDFETLEPEVSHETESHY-HVEETVSQDCNQDMSEMSQENPD 179  
 Db 896 EILPEDKNEKGOHEIVEE-ETLPDDDKNEKHEIVEEETL--PEDKNEKGOHEIVEV 952  
 QY 180 SSEPVDEHLHDDTDDTYQVYEQAVPEPLENEGIEITEVTAPEDNPVEDSQVIVEE 239  
 Db 953 EILPEDKNEKHEVE 967  
 QY 240 VSIFPVEEQVEPPD 254

## RESULT 9

ID Q20684 PRELIMINARY; PRT; 1262 AA.  
 AC Q20684;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
 DE F52H3.7 PROTEIN.  
 GN F52H3.7  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
 OC RHABDITINA; RHABDITOIDEA; RHABDITIIDAE; PELODERINAE; CAENORHABDITIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA GARDNER A.;  
 RA SUBMITTED (OCT-1995). TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94150718.  
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMLDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans."  
 RL NATURE 368:32-38(1994).  
 DR EMBL; 266512; E1351866;  
 SQ SEQUENCE 1262 AA; 133688 MW; 96ECBD5A CRC32;

Query Match 9.9%; Score 174; DB 5; Length 1262;

Best Local Similarity 21.6%; Pred. No. 2.55e-06;

Matches 49; Conservative 64; Mismatches 104; Indels 10; Gaps 8;

Db 654 DATTAADSSHEEPKRAESDAPAAESVETA-PVEEHKDEASPTALESVEDVIEAVHVASTES 712  
 QY 33 DADGDDFDVDDAKVLLGLKERSTSEPAVPPPEAPPHTEPEQVPEAPQNIIDEAKQ 92  
 Db 713 APAPVEDAASADSTOAPVEEQKDTAAAPESHEEAQAADHAASVEDVIEAVH--IASNESA 770  
 QY 93 IQSLLEHMHVHAHVEGE-DLQOEDGPTGEQOEDDEFLMATDVRDFETLEPEVSHETE 151  
 Db 771 PTTEAAPSESQAEDHKDEAATAEPAHEESVTVVEASHKEPQAESAA-PVEENIESVHV 829  
 QY 152 HSYHVEETVSQDCNQDMSEMSQENP-DSSEPVVEDERLHDDTDDTYQVYEE-QAVYE 209  
 Db 830 ASEESAPAPSGDAAPSEAPSVSENPV--EEVAVAVHVSIESVPTDA 874  
 QY 210 PLENEGIEITEVTAPEDNPVEDSQVIVEEVSIFPVEEQVEPPD 255

## RESULT 10

ID O13099 PRELIMINARY; PRT; 913 AA.  
 AC O13099;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE MIDDLE MOLECULAR WEIGHT NEUROFILAMENT PROTEIN NF-M.  
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;  
 OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA GERVASI C., SZARO B.G.;  
 RA BRAIN RES. MOL. BRAIN RES. 0:0-0(0).  
 CC -!- SIMILARITY: TO ALL OTHER INTERMEDIATE FILAMENT PROTEINS.  
 DR EMBL; U85970; G2062611;  
 DR PROSITE; PS00226; IF; 1.  
 DR PFAM; PF00038; Filament; 1.  
 KW INTERMEDIATE FILAMENT; COILED COIL; HEPTAD REPEAT PATTERN.  
 SQ SEQUENCE 913 AA; 102835 MW; 1B31AD72 CRC32;

Query Match 9.7%; Score 170; DB 13; Length 913;

Best Local Similarity 22.0%; Pred. No. 6.77e-06;

Matches 49; Conservative 56; Mismatches 111; Indels 7; Gaps 7;

Db 480 EETOGEDEEE-IVAIV-E-SSVQAAAPGEAEAGEEKEEKEEAEAE-EGE-KEE 534  
 QY 33 DADGDDFDVDDAKVLLGLKERSTSEPAVPPPEAPPHTEPEQVPEAPQNIIDEAKQ 92  
 Db 535 EKEEKEGEDEGEKVDEEEGEQGEEDVKAEKEEKEEKEEEDAEVETKAEKEEKEE 594  
 QY 93 IQSLLEHMHVHAHVEGEDLQOEDGPTGEQOEDDEFLMATDVRDFETLEPEVSHETE 152  
 Db 595 GEDAEKEEKEEKEEKEEVADEEGEKEEKEEKEEKEEKEEKEEKEEKEE-EGE-EGE 653  
 QY 153 SYHVEETVSQDCNQDMSEMSQENPDSEPVVEDERLHDDTDDTYQVYEEQAVYE 212  
 Db 654 EETTEAEAEVET-NTETKIVREKAEVQETAEETETA 695  
 QY 213 NEGIEITEVTAPEDNPVEDSQVIVEEVSIFPVEEQVEPPD 255



[illegible]

```
QY   202 YEEQAV--YEPLENEG-IETVETTAPPNPVEDSQSIVIEEV 240

RESULT      15
ID Q27043; PRELIMINARY; PRT; 543 AA.
AC Q27043;
DT DT 01-NOV-1996 (TREMBREL. 01, CREATED)
DT DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DE DE 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE DE POLYMORPHIC IMMUNODOMINANT MOLECULE.
GN GN PIM.
OS THEILERIA PARVA.
OC EUKARYOTA; ALVEOLATA; APICOPLASMA; PIROPLASMIDA; THEILERIADAE;
OC OC THEILERIA.
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN-ST7014;
RX RX MEDLINE: 95363102.
RA RA TOYE P.G., METZELAAR M.J., WIJNGAARD P.L., NENE V., IAMS K., ROOSE J.,
RA NYANJUJI J.K., GOBRIZLAIR E., MUSOKE A.J., CLEVERS H.C.;
RT RT "Characterization of the gene encoding the polymorphic immunodominant
RT molecule, a neutralizing antigen of Theileria parva.";
RL RL J. IMMUNOL.. 155:1370-1381(1995).
DR DR EMBL: U41833; G786136; -.
SR SR SEQUENCE 543 AA; 60026 MW; 68007BBD CRC32;

Query Match           8.6%; Score 152; DB 5; Length 543;
Best Local Similarity 22.2%; Pred.No. 4.98e-04;
Matches    44; Conservative     63; Mismatches 78; Indels 13; Gaps 12;

Db Db        62 DSSSTTSQQPQP-PDQPVDQQPPVVQSOPSGSDSPDPHQPTPDQPEDQG 120
   :|::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::||
Qy Qy       53 ERSTSEPAVPPEEAEPHTPEEQPVPEAPQNIDEAKIQISLLHEM-VHAHVESGDL 111

Db Db       121 QQQQ-QLDQTGGPG-LFPVDDQQGP-PQP-PQPERPQP-DQEDQGQ-QQLPDOQ 174
   ||:|| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Qy Qy      112 QLEDGTGEPOQEDDFLMATDVDDRFTLEPSHTEHSYHYEETSQCNDMEEM 171
   ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::||

Db Db       175 TGOOGQPIQPVDOQOVQEPAKDPTCG--QQQOPEPF-EPEPEOTPE-TTPQOEQPT 230
   ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::||
Qy Qy      172 MSEQENPDSEPVEVERLUHHDT-DDVTQVYEEQAYPELENEGIETEVTAPPEDNPV 230
   ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::||

Db Db       231 PDHPGSQQFQDPPEVGQ 248
   |:|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Qy Qy      231 ED SQVIVEVSIFPVEEQ 248

Search completed: Thu Oct 21 15:27:51 1999
Job time : 103 secs.
```

\*\*\*\*\*  
MORPH  
\*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:31:16 1999; MasPar time 3.42 Seconds  
Tabular output not generated. 62.103 Million cell updates/sec

Title: >US-09-040-485-6  
Description: (1-10) from US09040485.pep  
Perfect Score: 64  
Sequence: 1 APPEDNPVED 10  
Scoring table: PAM 150  
Gap 15

Searched: 170751 seqs, 21266608 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 14.620; Variance 50.711; scale 0.288

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	75.0	140	23	W19749	
2	47	73.4	129	37	W72935	Mouse inhibitor of ap 1.25e+02
3	46	71.9	20	32	W61358	Mycobacterium tubercu 1.56e+02
4	46	71.9	142	32	W61359	Survivin functionally 1.96e+02
5	45	70.3	797	35	W69373	Survivin protein. 1.96e+02
6	45	70.3	859	35	W69372	Modified retinoblasto 2.45e+02
7	45	70.3	869	35	W69374	Modified retinoblasto 2.45e+02
8	45	70.3	871	35	W69375	Modified retinoblasto 2.45e+02
9	45	70.3	897	35	W69371	Modified retinoblasto 2.45e+02
10	45	70.3	928	13	R71681	Recombinant ppl10RB p 2.45e+02
11	45	70.3	928	21	W09411	Retinoblastoma suscep 2.45e+02
12	45	70.3	928	13	R71680	Retinoblastoma p110R 2.45e+02
13	45	70.3	928	14	R74271	Retinoblastoma tumour 2.45e+02
14	45	70.3	928	1	R06289	Predicted retinoblast 2.45e+02
15	45	70.3	928	35	W69376	Modified retinoblasto 2.45e+02
16	45	70.3	928	7	R36534	Retinoblastoma (RB) p 2.45e+02

17	45	70.3	928	37	W62465	Retinoblastoma protei 2.45e+02
18	45	70.3	928	35	W69364	Modified retinoblasto 2.45e+02
19	45	70.3	928	35	W71354	Protein sequence of t 2.45e+02
20	45	70.3	928	1	R05305	Cancer supressing gen 2.45e+02
21	45	70.3	1028	34	W29667	Homo sapiens DL185_1 2.45e+02
22	45	70.3	2485	11	W59922	RAS associated GAP NF 2.45e+02
23	45	70.3	2485	11	R59921	RAS associated GAP NF 2.45e+02
24	45	70.3	2818	4	R22268	Nf1 gene product. 2.45e+02
25	45	70.3	2818	21	W13280	Human neurofibromin. 2.45e+02
26	44	68.8	395	16	R86553	Vibrio cholerae strai. 3.06e+02
27	44	68.8	395	16	R86555	Vibrio cholerae strai. 3.06e+02
28	44	68.8	437	2	R08390	Truncated form of hum 3.06e+02
29	44	68.8	491	15	R72866	Vibrio cholerae poly p 3.06e+02
30	44	68.8	828	2	R07661	Incomplete form of hu 3.06e+02
31	44	68.8	834	2	R08391	Sequence encoded by v 3.06e+02
32	44	68.8	979	2	R08338	Complete form of huma 3.06e+02
33	44	68.8	3413	30	W52849	A. mediterranei rifam 3.06e+02
34	43	67.2	100	38	W89015	Polypeptide fragment 3.81e+02
35	43	67.2	216	38	W89013	Polypeptide fragment 3.81e+02
36	43	67.2	510	4	R22365	SIVmac239 gag gene pr 3.81e+02
37	42	65.6	506	1	P80802	Sequence encoded by g 4.75e+02
38	42	65.6	521	1	P80807	Sequence of gag prote 4.75e+02
39	41	64.1	14	13	R64987	Amylopectin starch de 5.90e+02
40	41	64.1	360	13	R71677	Human CSBP2. 5.90e+02
41	41	64.1	403	23	W18084	Human Aurora-2. 5.90e+02
42	41	64.1	606	38	W85019	p38-green fluorescent 5.90e+02
43	41	64.1	607	38	W85010	p38-green fluorescent 5.90e+02
44	41	64.1	1189	13	R71463	Crystal protein encod 5.90e+02
45	41	64.1	1189	2	R10193	Insecticidal crystal 5.90e+02

ALIGNMENTS

RESULT 1  
ID W19749 standard; Protein; 140 AA.  
AC W19749;  
DT 16-SEP-1997 (first entry)  
DE Mouse inhibitor of apoptosis protein homologue MIHE.  
KW Inhibitor of apoptosis protein; IAP; mammalian IAP homologue; MIHE;  
KW degenerative disease; infectious disease; autoimmune disease;  
KW cancer; therapy; diagnosis.  
OS Mus musculus.  
PN W09723501-A1.  
PD 03-JUL-1997.  
PF 20-DEC-1996; AU0827.  
PR 22-DEC-1995; AU-007275.  
PA (AMRA-) AMRAD OPERATIONS PTY LTD.  
PI Vaux DL;  
DR WPI: 97-350966/32.  
DR N-PSDB; T72714.  
PT Isolated protein homologues of viral inhibitors of apoptosis - used  
PT to modulate apoptosis for treatment of degenerative, infectious or  
PT autoimmune diseases and cancer  
PS Claim 11: Page 71-72; 136pp; English.  
CC MIHE (W19749) is a murine homologue of baculovirus inhibitor of  
CC apoptosis protein (IAP). Its amino acid sequence was deduced from  
CC an isolated nucleic acid (see also T72714) obt'd. by a database  
CC search for sequences homologous to a baculovirus IAP repeat (BIR)  
CC consensus sequence (see also W19744). Unlike IAP, MIHE does not  
CC contain a RING finger domain. IAP homologues (see also W19745-48  
CC and W19750-52) and their derivatives and chemical analogues can be  
CC used in methods for modulating apoptosis in animal cells,  
CC specifically for treatment, by inhibition, of degenerative and  
CC infectious disease or, by promotion, of cancer and autoimmune  
CC disease.  
SQ Sequence 140 AA;

Query Match 75.0%; Score 48; DB 23; Length 140;  
Best Local Similarity 62.5%; Pred. No. 1.25e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 69 pdnpltee 76  
|:|:|:|:

```

QY      3 PEDNPVED 10

RESULT      2
ID W72935 standard; Protein; 129 AA.
AC W72935;
DE 21-JAN-1999 (first entry)
DT Mycobacterium tuberculosis antigen CFP8A.
KW Mycobacterium tuberculosis.
KW immunogen; infection.
OS Mycobacterium tuberculosis.
PN WO9844119-A1.
PD 08-OCT-1998.
PF 01-APR-1998; DK0132.
PR 03-JAN-1998; US-070488.
PR 02-APR-1997; DK-000376.
PR 18-APR-1997; US-044624.
PR 10-NOV-1997; DK-001277.
PA (STAT-) STATENS SERUM INST.
PI Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB,
PI Rosenkrands I, Welgeling K;
DR N-PSDB: V63945.
DR N-PSDB: V63945.
PT New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis.
PS Claim 1; Page 223; 163pp; English.
CC The present sequence represents a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments
CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis,
CC M. africanum or M. bovis.
SQ Sequence 129 AA;

Query Match      73.4%; Score 47; DB 37; Length 129;
Best Local Similarity 60.0%; Pred. No. 1.56e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db      28 apaganpvd 37
      ||: |||:|
QY      1 APPEDNPVED 10

RESULT      3
ID W61358 standard; peptide; 20 AA.
AC W61358;
DE 23-SEP-1998 (first entry)
DT Survivin functionally relevant peptide.
KW Survivin; apoptosis; cellular apoptosis; transplantaion;
KW motor neuron degenerative disease; HIV infection; immunosuppression;
KW gastrointestinal perturbations; cardiovascular disorder.
OS Homo sapiens.
PN WO9822589-A2.
PD 28-MAY-1998.
PF 20-NOV-1997; U21880.
PR 20-NOV-1997; US-975080.
PR 20-NOV-1996; US-031435.
PA (UYIA ) UNIV YALE.
PI Altieri DC;
DR N-PSDB: V27941.
PT Modulating apoptosis by controlling the Survivin gene - useful for
PT treating transplant rejection, degenerative disorders and tumours
PS Disclosure; Fig 10; 108pp; English.
CC The survivin gene can be used to control apoptosis through modification
CC of the gene. Survivin peptides can be used to inhibit cellular
CC apoptosis, e.g. for enhancing the viability of organs and tissues prior
CC to their transplantation, for preserving the growth of cells in culture
CC or for treating conditions involving abnormal apoptosis,
CC e.g. degenerative diseases such as motor neuron degenerative diseases,
CC HIV infection, dermatological effects of ageing, disorders and diseases
CC such as immunosuppression, gastrointestinal perturbations, cardiovascular
CC disorders, apoptosis related to reperfusion damage, rejection of tissue
CC transplantation and Alzheimer's disease. Agents which block Survivin
CC activity can be used to treat e.g. tumours.
SQ Sequence 142 AA;

Query Match      71.9%; Score 46; DB 32; Length 142;
Best Local Similarity 50.0%; Pred. No. 1.96e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db      69 pdddpiece 76
      ||: |||:|
QY      3 PEDNPVED 10

RESULT      5
ID W69373 standard; Protein; 797 AA.
AC W69373;
DE 02-DEC-1998 (first entry)
DT Modified retinoblastoma tumour suppressor.
KW Modified retinoblastoma tumour suppressor;
KW cellular proliferation inhibitor.
OS Homo sapiens.
PN WO9837091-A2.
PD 27-AUG-1998.
PF 19-FEB-1998; U03041.
PR 20-FEB-1997; US-038118.
PA (BAYU ) BAYLOR COLLEGE MEDICINE.

```

e.g. degenerative diseases such as motor neuron degenerative diseases, HIV infection, dermatological effects of ageing, disorders and diseases such as immunosuppression, gastrointestinal perturbations, cardiovascular disorders, apoptosis related to reperfusion damage, rejection of tissue transplantation and Alzheimer's disease. Agents which block Survivin activity can be used to treat e.g. tumours.

Query Match 71.9%; Score 46; DB 32; Length 20; Best Local Similarity 50.0%; Pred. No. 1.96e+02; Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 5 pdddpiece 12  
||: |||:|  
QY 3 PEDNPVED 10

RESULT 4  
ID W61359 standard; Protein; 142 AA.  
AC W61359;  
DE 25-SEP-1998 (first entry)  
DT Survivin protein.  
KW survivin; apoptosis; cellular apoptosis; transplantaion;  
KW motor neuron degenerative disease; HIV infection; immunosuppression;  
KW gastrointestinal perturbations; cardiovascular disorder.  
OS Homo sapiens.  
PN WO9822589-A2.  
PD 28-MAY-1998.  
PF 20-NOV-1997; U21880.  
PR 20-NOV-1997; US-975080.  
PR 20-NOV-1996; US-031435.  
PA (UYIA ) UNIV YALE.  
PI Altieri DC;  
DR WPI; 98-312475/27.  
DR N-PSDB: V27941.  
PT Modulating apoptosis by controlling the Survivin gene - useful for  
PT treating transplant rejection, degenerative disorders and tumours  
PS Disclosure; Fig 10; 108pp; English.  
CC The survivin gene can be used to control apoptosis through modification  
CC of the gene. Survivin peptides can be used to inhibit cellular  
CC apoptosis, e.g. for enhancing the viability of organs and tissues prior  
CC to their transplantation, for preserving the growth of cells in culture  
CC or for treating conditions involving abnormal apoptosis,  
CC e.g. degenerative diseases such as motor neuron degenerative diseases,  
CC HIV infection, dermatological effects of ageing, disorders and diseases  
CC such as immunosuppression, gastrointestinal perturbations, cardiovascular  
CC disorders, apoptosis related to reperfusion damage, rejection of tissue  
CC transplantation and Alzheimer's disease. Agents which block Survivin  
CC activity can be used to treat e.g. tumours.  
SQ Sequence 142 AA;

Query Match 71.9%; Score 46; DB 32; Length 142; Best Local Similarity 50.0%; Pred. No. 1.96e+02; Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 69 pdddpiece 76  
||: |||:|  
QY 3 PEDNPVED 10

RESULT 5  
ID W69373 standard; Protein; 797 AA.  
AC W69373;  
DE 02-DEC-1998 (first entry)  
DT Modified retinoblastoma tumour suppressor.  
KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;  
KW cellular proliferation inhibitor.  
OS Homo sapiens.  
PN WO9837091-A2.  
PD 27-AUG-1998.  
PF 19-FEB-1998; U03041.  
PR 20-FEB-1997; US-038118.  
PA (BAYU ) BAYLOR COLLEGE MEDICINE.

(TEXA ) UNIV TEXAS SYSTEM.  
 PA Benedict WF, Hu S, Xu H, Zhou Y;  
 PI WPI: 98-480788/41.  
 DR N-PSDB: V58449.  
 PT Retinoblastoma suppressor protein with N-terminal modification -  
 PT Inhibiting cellular proliferation, particularly cancer  
 PS Claim 21; Page 203-207; 249pp; English.  
 CC This sequence represents a modified retinoblastoma tumour suppressor  
 CC protein (RTSP) of the invention. The proteins can be used for inhibiting  
 CC cellular proliferation, when coadministered with a p53 protein. The RTSPs  
 CC can be used for treating diseases characterised by abnormal cellular  
 CC proliferation, particularly cancers. The RTSPs have a broader spectrum of  
 CC activity than wild type RTSPs.  
 SQ Sequence 797 AA;

Query Match 70.3%; Score 45; DB 35; Length 797;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 ppspeedpeqd 36  
 :||||:|:  
 QY 1 APPEDNPVED 10

## RESULT 6

ID W69372 standard; Protein; 859 AA.  
 AC W69372;  
 DT 02-DEC-1998 (first entry)  
 DE Modified retinoblastoma tumour suppressor.  
 KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;  
 KW cellular proliferation inhibitor.  
 OS Homo sapiens.  
 PN W09837091-A2.  
 PD 27-AUG-1998.  
 PF 19-FEB-1998; U03041.  
 PR 20-FEB-1997; US-038118.  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PI (TEXA ) UNIV TEXAS SYSTEM.  
 PI Benedict WF, Hu S, Xu H, Zhou Y;  
 DR WPI: 98-480788/41.  
 DR N-PSDB: V58448.  
 PT Retinoblastoma suppressor protein with N-terminal modification -  
 PT Inhibiting cellular proliferation, particularly cancer  
 PS Claim 21; Page 196-200; 249pp; English.  
 CC This sequence represents a modified retinoblastoma tumour suppressor  
 CC protein (RTSP) of the invention. The proteins can be used for inhibiting  
 CC cellular proliferation, when coadministered with a p53 protein. The RTSPs  
 CC can be used for treating diseases characterised by abnormal cellular  
 CC proliferation, particularly cancers. The RTSPs have a broader spectrum of  
 CC activity than wild type RTSPs.  
 SQ Sequence 859 AA;

Query Match 70.3%; Score 45; DB 35; Length 859;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 ppspeedpeqd 36  
 :||||:|:  
 QY 1 APPEDNPVED 10

## RESULT 7

ID W69374 standard; Protein; 869 AA.  
 AC W69374;  
 DT 02-DEC-1998 (first entry)  
 DE Modified retinoblastoma tumour suppressor.  
 KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;  
 KW cellular proliferation inhibitor.  
 OS Homo sapiens.  
 PN W09837091-A2.  
 PD 27-AUG-1998.  
 PF 19-FEB-1998; U03041.  
 PR 20-FEB-1997; US-038118.

(BAYU ) BAYLOR COLLEGE MEDICINE.  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PI Benedict WF, Hu S, Xu H, Zhou Y;  
 DR WPI: 98-480788/41.  
 DR N-PSDB: V58450.  
 PT Retinoblastoma suppressor protein with N-terminal modification -  
 PT Inhibiting cellular proliferation, particularly cancer  
 PS Claim 21; Page 210-215; 249pp; English.  
 CC This sequence represents a modified retinoblastoma tumour suppressor  
 CC protein (RTSP) of the invention. The proteins can be used for inhibiting  
 CC cellular proliferation, when coadministered with a p53 protein. The RTSPs  
 CC can be used for treating diseases characterised by abnormal cellular  
 CC proliferation, particularly cancers. The RTSPs have a broader spectrum of  
 CC activity than wild type RTSPs.  
 SQ Sequence 869 AA;

Query Match 70.3%; Score 45; DB 35; Length 869;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 ppspeedpeqd 36  
 :||||:|:  
 QY 1 APPEDNPVED 10

## RESULT 8

ID W69375 standard; Protein; 871 AA.  
 AC W69375;  
 DT 02-DEC-1998 (first entry)  
 DE Modified retinoblastoma tumour suppressor.  
 KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;  
 KW cellular proliferation inhibitor.  
 OS Homo sapiens.  
 PN W09837091-A2.  
 PD 27-AUG-1998.  
 PF 19-FEB-1998; U03041.  
 PR 20-FEB-1997; US-038118.  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PI (TEXA ) UNIV TEXAS SYSTEM.  
 PI Benedict WF, Hu S, Xu H, Zhou Y;  
 DR WPI: 98-480788/41.  
 DR N-PSDB: V58451.  
 PT Retinoblastoma suppressor protein with N-terminal modification -  
 PT Inhibiting cellular proliferation, particularly cancer  
 PS Claim 21; Page 218-222; 249pp; English.  
 CC This sequence represents a modified retinoblastoma tumour suppressor  
 CC protein (RTSP) of the invention. The proteins can be used for inhibiting  
 CC cellular proliferation, when coadministered with a p53 protein. The RTSPs  
 CC can be used for treating diseases characterised by abnormal cellular  
 CC proliferation, particularly cancers. The RTSPs have a broader spectrum of  
 CC activity than wild type RTSPs.  
 SQ Sequence 871 AA;

Query Match 70.3%; Score 45; DB 35; Length 871;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 ppspeedpeqd 36  
 :||||:|:  
 QY 1 APPEDNPVED 10

## RESULT 9

ID W69371 standard; Protein; 897 AA.  
 AC W69371;  
 DT 02-DEC-1998 (first entry)  
 DE Modified retinoblastoma tumour suppressor.  
 KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;  
 KW cellular proliferation inhibitor.  
 OS Homo sapiens.  
 PN W09837091-A2.  
 PD 27-AUG-1998.  
 PF 19-FEB-1998; U03041.

PR 20-FEB-1997; US-038118.  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PI (TEXA ) UNIV TEXAS SYSTEM  
 DR Benedict WF, Hu S, Xu H, Zhou Y;  
 DR WPI: 98-480788/41.  
 DR N-PSDB; V58447.  
 PT Retinoblastoma suppressor protein with N-terminal modification -  
 PT inhibiting cellular proliferation, particularly cancer  
 PS Claim 21; Page 188-192; 249pp; English.  
 CC This sequence represents a modified retinoblastoma tumour suppressor  
 CC protein (R1SP) of the invention. The proteins can be used for inhibiting  
 CC cellular proliferation, when coadministered with a p53 protein. The R1SPs  
 CC can be used for treating diseases characterised by abnormal cellular  
 CC proliferation, particularly cancers. The R1SPs have a broader spectrum of  
 CC activity than wild type R1SPs.  
 SQ Sequence 897 AA;

Query Match 70.3%; Score 45; DB 35; Length 897;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppdeeqd 36  
 :|||:|:|  
 QY 1 APPEDNPVED 10

RESULT 10  
 ID R71681 standard; Protein; 928 AA.  
 AC R71681;  
 DT 16-OCT-1995 (first entry)  
 DE Recombinant p110RB protein.  
 KW Rb110; retinoblastoma; cancer; tumor suppressor protein;  
 KW cell cycle; p110RB; Escherichia coli.  
 OS Homo sapiens.  
 FH key Location/Qualifiers  
 FT misc\_difference 2  
 FT /note= "amino acid at position 2 is Pro in the  
 FT natural sequence"  
 FT  
 PN W09507708-A.  
 PD 23-MAR-1995.  
 PR 13-SEP-1994; U10357.  
 PR 13-SEP-1993; US-121108.  
 PA (CANJ-) CANJI INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Goodrich DW, Johnson D, Lee EYP, Lee W, Shepard HM;  
 PI Wang NP;  
 DR WPI: 95-131179/17.  
 PT Admin. of a functional retinoblastoma polypeptide or protein -  
 PT used to prevent and inhibit prim. and sec. retinoblastoma-linked  
 PT cancers  
 PS Disclosure; Fig.30; 163pp; English.  
 CC By chromosomal walking from the esterase D gene on chromosome 13,  
 CC the retinoblastoma susceptibility gene (RB) was identified on the  
 CC basis of chromosomal location, homologous deletion and tumor-specific  
 CC alterations in expression. Screening of cDNA libraries yielded clone  
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and  
 CC encoding a 110-140 kDa nuclear phosphoprotein named p110RB (R71680).  
 CC Recombinant p110RB was prepared in E. coli, but with a P2A mutation  
 CC for cloning convenience (R71681).  
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 13; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppdeeqd 36  
 :|||:|:|  
 QY 1 APPEDNPVED 10

RESULT 11  
 ID W09411 standard; Protein; 928 AA.  
 AC W09411;

DT 22-JUN-1997 (first entry)  
 DE Retinoblastoma susceptibility phosphoprotein p110RB.  
 KW Retinoblastoma susceptibility gene; p110RB; nuclear phosphoprotein;  
 KW cancer; osteosarcoma; fibrosarcoma; glioblastoma; breast cancer;  
 KW polyclonal antibody; diagnosis.  
 OS Homo sapiens.  
 PN US5578701-A.  
 PD 26-NOV-1996.  
 PF 17-SEP-1987; 098612.  
 PR 11-SEP-1987; US-098612.  
 PR 11-JUL-1990; US-550877.  
 PR 14-JUL-1992; US-914039.  
 PR 17-JUN-1993; US-079207.  
 PR 08-APR-1994; US-225099.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Lee EYP, Lee W;  
 PI WPI: 97-020465/02.  
 DR Retinoblastoma phosphoprotein p110RB-specific polyclonal antibody  
 PT - for diagnosing retinoblastoma and other related tumours or  
 PT susceptibility to them  
 PS Disclosure; Fig 2; 21pp; English.  
 CC A 110-114 kDa phosphoprotein product (W09411) of the retinoblastoma  
 CC susceptibility gene is designated p110RB. Its sequence was deduced  
 CC from a full-length cDNA sequence that included the sequence of a  
 CC clone derived from the retinoblastoma cell line V79. p110RB is  
 CC primarily located in the cell nucleus and has DNA binding activity.  
 CC The absence or loss of p110RB mediates oncogenicity. p110RB-  
 CC specific polyclonal antibodies can be utilised in methods of  
 CC diagnosing hereditary predisposition to retinoblastoma or to other  
 CC diseases controlled by the retinoblastoma gene such as  
 CC osteosarcoma, fibrosarcoma, glioblastoma and breast cancer.  
 CC Provision of p110RB to an individual through molecular induction  
 CC and gene transplanting may be used as a means of suppressing  
 CC tumorigenesis.  
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 21; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppdeeqd 36  
 :|||:|:|  
 QY 1 APPEDNPVED 10

RESULT 12  
 ID R71680 standard; Protein; 928 AA.  
 AC R71680;  
 DT 16-OCT-1995 (first entry)  
 DE Retinoblastoma p110RB protein.  
 KW Rb110; retinoblastoma; cancer; tumor suppressor protein;  
 KW cell cycle; p110RB.  
 OS Homo sapiens.  
 PN W09507708-A.  
 PD 23-MAR-1995.  
 PF 13-SEP-1994; U10357.  
 PR 13-SEP-1993; US-121108.  
 PA (CANJ-) CANJI INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Goodrich DW, Johnson D, Lee EYP, Lee W, Shepard HM;  
 PI Wang NP;  
 DR WPI: 95-131179/17.  
 DR N-PSDB; Q86398.  
 PT Admin. of a functional retinoblastoma polypeptide or protein -  
 PT used to prevent and inhibit prim. and sec. retinoblastoma-linked  
 PT cancers  
 PS Disclosure; Fig.2; 163pp; English.  
 CC By chromosomal walking from the esterase D gene on chromosome 13,  
 CC the retinoblastoma susceptibility gene (RB) was identified on the  
 CC basis of chromosomal location, homologous deletion and tumor-specific  
 CC alterations in expression. Screening of cDNA libraries yielded clone  
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and  
 CC encoding a 110-140 kDa nuclear phosphoprotein named p110RB (R71680).  
 CC

CC Use of recombinant p110RB will reduce the need for conventional  
 CC radiotherapy or chemotherapy.

SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 13; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppeedpeqd 36  
 :||:|:|  
 QY 1 APPEDNPVED 10

#### RESULT 13

ID R74271 standard; Protein; 928 AA.  
 AC R74271;  
 DT 28-NOV-1995 (first entry)  
 DE Retinoblastoma tumour suppressor protein.  
 KW Recombinant; adenovirus; expression vector; TSG; small lung cancer;  
 KW hepatocarcinoma; melanoma; retinoblastoma; sarcoma; sickle cell;  
 KW anaemia; Tay-Sach's disease.  
 OS Homo sapiens.  
 PN W09511984-A.  
 PD 04-MAY-1995.  
 PF 25-OCT-1994; U12235.  
 PR 25-OCT-1993; US-142669.  
 PR 19-MAY-1994; US-246007.  
 PA (CANJ-) CANJ INC.  
 PI Gregory RJ, Maneval DC, Wills KN;  
 DR WPI; 95-178976/23.  
 DR N-PSDB; Q90059.  
 PT Adenoviral vector with deletion of viral protein IX contains  
 foreign gene - esp. encoding tumour suppressor protein for gene  
 therapy of tumours, reduces contamination by wild type adenovirus  
 PS Disclosure; Fig 3; 92pp; English.  
 CC The sequence is that of a retinoblastoma tumour protein. The gene  
 encoding this protein may be used in a novel method involving a  
 CC recombinant adenovirus expression vector to treat diseases  
 CC associated with the absence of the TSG or the presence of a mutated  
 CC TSG, e.g. many forms of carcinoma, sickle cell anaemia or Tay-Sach's  
 CC disease.  
 CC See also R74272.  
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 14; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppeedpeqd 36  
 :||:|:|  
 QY 1 APPEDNPVED 10

#### RESULT 14

ID R06289 standard; protein; 928 AA.  
 AC R06289;  
 DT 13-DEC-1990 (first entry)  
 DE Predicted retinoblastoma gene product.  
 KW Osteosarcoma; fibrosarcoma; glioblastoma; breast cancer; ds.  
 OS Homo sapiens.  
 PN US4942123-A.  
 PD 17-JUL-1990.  
 PF 17-SEP-1987; 098612.  
 PR 17-SEP-1987; US-098612.  
 PA (REGC ) UNIV OF CALIFORNIA.  
 PI Lee WH, Eva Y, Lee HP;  
 DR WPI; 90-245977/32.  
 PT Diagnosing absence or inactivation of retinoblastoma gene - by  
 PT detecting the absence of specific anti-p107 antibody  
 PT immuno-complex formed using tissue  
 PS Disclosure; p; English.  
 CC Laelled Abs raised to the RB gene product may be used to screen  
 CC for RB and in diagnosis of susceptibility to associated secondary

CC cancers such as osteosarcoma, fibrosarcoma, glioblastoma and  
 CC breast cancer.

SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 1; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppeedpeqd 36  
 :||:|:|  
 QY 1 APPEDNPVED 10

#### RESULT 15

ID W69376 standard; Protein; 928 AA.  
 AC W69376;  
 DT 02-DEC-1998 (first entry)  
 DE Modified retinoblastoma tumour suppressor.  
 KW Modified retinoblastoma tumour suppressor; RTSP protein; cancer therapy;  
 KW cellular proliferation inhibitor.  
 OS Homo sapiens.  
 PN W09837091-A2.  
 PD 27-AUG-1998.  
 PF 19-FEB-1998; U03041.  
 PR 20-FEB-1997; US-038118.  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PI Benedict WF, Hu S, Xu H, Zhou Y;  
 DR WPI; 98-480788/41.  
 DR N-PSDB; V58452.  
 PT Retinoblastoma suppressor protein with N-terminal modification -  
 PT inhibiting cellular proliferation, particularly cancer  
 PS Claim 21; Page 226-230; 249pp; English.  
 CC This sequence represents a modified retinoblastoma tumour suppressor  
 CC protein (RTSP) of the invention. The proteins can be used for inhibiting  
 CC cellular proliferation, when coadministered with a p53 protein. The RTSPs  
 CC can be used for treating diseases characterised by abnormal cellular  
 CC proliferation, particularly cancers. The RTSPs have a broader spectrum of  
 CC activity than wild type RTSPs.  
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 35; Length 928;  
 Best Local Similarity 50.0%; Pred. No. 2.45e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 pppeedpeqd 36  
 :||:|:|  
 QY 1 APPEDNPVED 10

Search completed: Thu Oct 21 15:31:35 1999  
 Job time : 19 secs.

\*\*\*\*\*

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Thu Oct 21 15:33:50 1999; MasPar time 1.52 Seconds  
Tabular output not generated. 76.957 Million cell updates/sec

Title: >US-09-040-485-6  
Description: (1-10) from US09040485.pep  
Perfect Score: 64  
Sequence: 1 APPEDNPVED 10

Scoring table: PAM 150  
Gap 15

Searched: 119857 seqs, 11713122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 13.702; Variance 48.684; scale 0.281

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	70.3	365	2	US-08-855- Sequence 5, Applicatio	1.31e+02
2	45	70.3	928	3	PCT-US94-1 Sequence 3, Applicatio	1.31e+02
3	45	70.3	928	3	PCT-US94-1 Sequence 2, Applicatio	1.31e+02
4	45	70.3	928	2	US-08-959- Sequence 8, Applicatio	1.31e+02
5	45	70.3	928	1	US-08-204- Sequence 1, Applicatio	1.31e+02
6	45	70.3	2020	1	US-07-551- Sequence 2, Applicatio	1.31e+02
7	45	70.3	2485	3	PCT-US94-0 Sequence 2, Applicatio	1.31e+02
8	45	70.3	2485	3	PCT-US94-0 Sequence 1, Applicatio	1.31e+02
9	45	70.3	2818	1	US-08-510- Sequence 1, Applicatio	1.31e+02
10	45	70.3	2818	1	US-08-411- Sequence 2, Applicatio	1.31e+02
11	45	70.3	2818	2	US-08-449- Sequence 2, Applicatio	1.31e+02
12	44	68.8	36	1	US-07-776- Sequence 18, Applicatio	1.64e+02
13	44	68.8	395	1	US-07-931- Sequence 2, Applicatio	1.64e+02
14	44	68.8	395	2	US-08-624- Sequence 15, Applicatio	1.64e+02
15	44	68.8	395	1	US-07-931- Sequence 5, Applicatio	1.64e+02
16	42	65.6	652	2	US-08-313- Sequence 53, Applicatio	2.55e+02
17	42	65.6	652	2	US-08-459- Sequence 17, Applicatio	2.55e+02
18	41	64.1	360	3	PCT-US94-1 Sequence 12, Applicatio	3.18e+02
19	41	64.1	360	2	US-08-950- Sequence 12, Applicatio	3.18e+02
20	41	64.1	360	3	PCT-US94-1 Sequence 14, Applicatio	3.18e+02
21	41	64.1	360	1	US-08-674- Sequence 3, Applicatio	3.18e+02
22	41	64.1	360	2	US-08-605- Sequence 12, Applicatio	3.18e+02
23	41	64.1	360	2	US-08-950- Sequence 14, Applicatio	3.18e+02

24	41	64.1	360	2	US-08-605- Sequence 14, Applicati	3.18e+02
25	41	64.1	360	2	US-08-469- Sequence 12, Applicati	3.18e+02
26	41	64.1	360	2	US-08-920- Sequence 3, Applicati	3.18e+02
27	41	64.1	360	2	US-08-250- Sequence 12, Applicati	3.18e+02
28	41	64.1	360	2	US-08-250- Sequence 5, Applicati	3.18e+02
29	41	64.1	520	3	PCT-US95-0 Sequence 2, Applicatio	3.18e+02
30	41	64.1	823	3	US-08-461- Sequence 2, Applicatio	3.18e+02
31	41	64.1	1176	1	US-08-457- Sequence 2, Applicatio	3.18e+02
32	41	64.1	1189	4	5188960-6 Patent No. 5188960	3.18e+02
33	41	64.1	1189	2	US-08-757- Sequence 6, Applicatio	3.18e+02
34	41	64.1	1189	2	US-08-980- Sequence 2, Applicatio	3.18e+02
35	41	64.1	1189	2	US-08-757- Sequence 10, Applicati	3.18e+02
36	41	64.1	1189	1	US-08-602- Sequence 2, Applicatio	3.18e+02
37	41	64.1	1189	2	US-08-757- Sequence 12, Applicati	3.18e+02
38	41	64.1	1189	2	US-08-980- Sequence 12, Applicati	3.18e+02
39	41	64.1	1189	2	US-08-757- Sequence 4, Applicatio	3.18e+02
40	41	64.1	1189	3	PCT-US92-1 Sequence 16, Applicati	3.18e+02
41	41	64.1	1189	2	US-08-757- Sequence 8, Applicatio	3.18e+02
42	41	64.1	1189	2	US-08-980- Sequence 8, Applicatio	3.18e+02
43	41	64.1	1189	2	US-08-980- Sequence 61, Applicati	3.18e+02
44	41	64.1	1190	2	US-08-732- Sequence 12, Applicati	3.18e+02
45	41	64.1	1190	1	US-08-239- Sequence 12, Applicati	3.18e+02

ALIGNMENTS

RESULT 1  
ID US-08-855-518-5 STANDARD; PRT; 365 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
Sequence 5, Application US/08855518  
XX  
CC Patent No. 5858675  
CC Sequence 5, Application US/08855518  
CC GENERAL INFORMATION:  
CC APPLICANT: Hillman, Jennifer L.  
CC APPLICANT: Lal, Preeti  
CC APPLICANT: Neil C. Corley  
CC TITLE OF INVENTION: DOUBLE-STRANDED RNA-BINDING  
CC TITLE OF INVENTION: PROTEIN  
CC NUMBER OF SEQUENCES: 5  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.  
CC STREET: 3174 Porter Drive  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94304  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FastSeq for Windows Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/855,518  
CC FILING DATE: Filed Herewith  
CC CLASSIFICATION: 514  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Billings, Lucy J.  
CC REGISTRATION NUMBER: 36,749  
CC REFERENCE/DOCKET NUMBER: PF-0299 US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 415-855-0555  
CC TELEFAX: 415-845-4166  
CC INFORMATION FOR SEQ ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 365 amino acids





CC TITLE OF INVENTION: Recombinant Adenoviral Vector and  
CC TITLE OF INVENTION: Methods of Use  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Campbell and Flores  
CC STREET: 4370 La Jolla Village Drive, Suite 700  
CC CITY: San Diego  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 92122  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/959,638  
CC FILING DATE:  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/328,673  
CC FILING DATE: 25-OCT-1994  
CC APPLICATION NUMBER: US 08/233,777  
CC FILING DATE: 19-MAY-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/142,669  
CC FILING DATE: 25-OCT-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Campbell, Cathryn A.  
CC REGISTRATION NUMBER: 31,815  
CC REFERENCE/DOCKET NUMBER: P-CJ 1192  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (619) 535-9001  
CC TELEFAX: (619) 535-8949  
CC INFORMATION FOR SEQ ID NO: 8:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 928 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 928 AA; 106158 MW; 4596348 CN;  
SQ  
Query Match 70.38; Score 45; DB 2; Length 928;  
Best Local Similarity 50.08; Pred. No. 1.31e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 27 PPPEDEPQD 36  
QY 1 APPEDNPVED 10  
RESULT 5  
ID US-08-204-329-1 STANDARD; PRT; 928 AA.  
XX  
AC xxxxxx  
XX  
DT  
DE  
Sequence 1, Application US/08204329  
XX  
Sequence 1, Application US/08204329  
CC Patent No. 5710255  
CC GENERAL INFORMATION:  
CC APPLICANT: SHEPARD, H. M.  
CC APPLICANT: WEN, SHU F.  
CC TITLE OF INVENTION: CHARACTERIZATION OF A NOVEL ANTI-P110RB  
CC NUMBER OF SEQUENCES: 2  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP  
CC STREET: TWO EMBARCADERO CENTER, 8TH FLOOR  
CC CITY: SAN FRANCISCO  
CC STATE: CALIFORNIA

CC COUNTRY: U.S.A.  
CC ZIP: 94111  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/204,329  
CC FILING DATE: 15-AUG-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US92/05866  
CC FILING DATE: 14-JUL-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: RENEE A. FITTS  
CC REGISTRATION NUMBER: 35,136  
CC REFERENCE/DOCKET NUMBER: 16930-00040005  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 326-2400  
CC TELEFAX: (415) 326-2422  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 928 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC SEQUENCE 928 AA; 106158 MW; 4596348 CN;  
SQ  
Query Match 70.38; Score 45; DB 1; Length 928;  
Best Local Similarity 50.08; Pred. No. 1.31e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 27 PPPEDEPQD 36  
QY 1 APPEDNPVED 10  
RESULT 6  
ID US-07-551-531-2 STANDARD; PRT; 2020 AA.  
XX  
AC xxxxxx  
XX  
DT  
DE  
Sequence 2, Application US/07551531  
XX  
Sequence 2, Application US/07551531  
CC Patent No. 5227292  
CC GENERAL INFORMATION:  
CC APPLICANT: WHITE, Raymond L.  
CC APPLICANT: O'CONNELL, Peter  
CC APPLICANT: VISKOCHEL, David H.  
CC APPLICANT: CAWTHON, Richard M.  
CC TITLE OF INVENTION: NEUROFIBROMATOSIS TYPE 1 GENE  
CC NUMBER OF SEQUENCES: 2  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Venable, Baetjer, Howard & Civiletti  
CC STREET: 1201 New York Avenue, N.W., Suite 1000  
CC CITY: Washington  
CC STATE: DC  
CC ZIP: 20005  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/551,531  
CC FILING DATE: 19900712  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:

CC NAME: IHNNEN, Jeffrey L.  
CC REGISTRATION NUMBER: 28,957  
CC REFERENCE/DOCKET NUMBER: 19780-94649  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202-962-4810  
CC TELEFAX: 202-962-8300  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2020 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 2020 AA; 226858 MW; 21863010 CN;

Query Match 70.3%; Score 45; DB 1; Length 2020;  
Best Local Similarity 60.0%; Pred. No. 1.31e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 728 GPPEHKPVAD 737  
QY 1 APPEDNPVED 10

RESULT 7  
ID PCT-US94-00198-2 STANDARD; PRT; 2485 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
XX  
DE  
XX  
XX

Sequence 2, Application PC/TUS9400198

Sequence 2, Application PC/TUS9400198  
GENERAL INFORMATION:  
APPLICANT: Schering Corp.  
TITLE OF INVENTION: RAS Associated GAP Proteins  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schering Corp.  
STREET: 1 Girald Farms  
CITY: Madison  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 94304-1104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: 6.0.8  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/00198  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/004,824  
FILING DATE: 15-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Paul G.  
REGISTRATION NUMBER: 32,743  
REFERENCE/DOCKET NUMBER: DX0352 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201)822-7255  
TELEFAX: (201)822-7039  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2485 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
SEQUENCE 2485 AA; 279289 MW; 32942281 CN;

Query Match 70.3%; Score 45; DB 3; Length 2485;  
Best Local Similarity 60.0%; Pred. No. 1.31e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1192 GPPEHKPVAD 1201  
QY 1 APPEDNPVED 10

RESULT 8  
ID PCT-US94-00198-1 STANDARD; PRT; 2485 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
XX  
DE  
XX  
XX

Sequence 1, Application PC/TUS9400198

Sequence 1, Application PC/TUS9400198  
GENERAL INFORMATION:  
APPLICANT: Schering Corp.  
TITLE OF INVENTION: RAS Associated GAP Proteins  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schering Corp.  
STREET: 1 Girald Farms  
CITY: Madison  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: 6.0.8  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/00198  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/004,824  
FILING DATE: 15-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Paul G.  
REGISTRATION NUMBER: 32,743  
REFERENCE/DOCKET NUMBER: DX0352 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201)822-7255  
TELEFAX: (201)822-7039  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2485 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 564..9380  
SEQUENCE 2485 AA; 279325 MW; 32915779 CN;

Query Match 70.3%; Score 45; DB 3; Length 2485;  
Best Local Similarity 60.0%; Pred. No. 1.31e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1192 GPPEHKPVAD 1201  
QY 1 APPEDNPVED 10

```

RESULT 9
ID US-08-510-284-1 STANDARD; PRT; 2818 AA.
XX
AC xxxxxx
XX
XX
DT
XX
XX
Sequence 1, Application US/08510284
XX
Sequence 1, Application US/08510284
XX
Patent No. 5580955
CC
GENERAL INFORMATION:
CC
APPLICANT: Nur-E-kamal, M. S. A.; Maruta, Hiroshi
CC
TITLE OF INVENTION: FRAGMENTS OF NEUROFIBROMIN (NFI) AND METHOD
CC
TITLE OF INVENTION: TO REVERSE ACTIVATED RAS INDUCED MALIGNANT TRANSFORMATION
CC
TITLE OF INVENTION: MAMMALIAN CELLS
CC
NUMBER OF SEQUENCES: 2
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: Felfe & Lynch
CC
STREET: 805 Third Avenue
CC
CITY: New York City
CC
STATE: New York
CC
COUNTRY: USA
CC
ZIP: 10022
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
CC
COMPUTER: IBM PS/2
CC
OPERATING SYSTEM: PC-DOS
CC
SOFTWARE: Wordperfect
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/510,284
CC
FILING DATE:
CC
CLASSIFICATION: 514
CC
PRIOR APPLICATION DATA:
CC
APPLICATION NUMBER: 08/071,575
CC
FILING DATE: 1-JUNE-1993
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Hanson, No. 5580955man D.
CC
REGISTRATION NUMBER: 30,946
CC
REFERENCE/DOCKET NUMBER: LUD 5319
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: (212) 688-9200
CC
TELEFAX: (212) 838-3884
CC
INFORMATION FOR SEQ ID NO: 1:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 2818 amino acids
CC
TYPE: amino acid
CC
TOPOLOGY: linear
CC
MOLECULE TYPE: protein
CC
PUBLICATION INFORMATION:
CC
AUTHORS: Marchuk, Douglas A.; Saulino, Ann M.;
CC
AUTHORS: Tavakoli, Roxanne; Swaroop, Manju;
CC
AUTHORS: Wallace, Margaret R.; Andersen, Lone B.;
CC
AUTHORS: Mitchell, Anna L.; Gutmann, David H.;
CC
AUTHORS: Boguski, Mark; Collins, Francis S.
CC
TITLE: cDNA Cloning of the Type 1 Neurofibromatosis Gene:
CC
TITLE: Complete Sequence of the NF1 Gene Product
CC
Patent No. 5580955
CC
JOURNAL: Genomics
CC
VOLUME: 11
CC
PAGES: 931-940
CC
DATE: 1991
CC
SEQUENCE 2818 AA; 317030 MW; 42201960 CN;
SQ
Query Match 70.3%; Score 45; DB 1; Length 2818;
Best Local Similarity 60.0%; Pred. No. 1.31e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1526 GPEKHPVAD 1535
Oy 1 APPENDPVED 10

```

```

RESULT 10
XX US-08-411-389-2 STANDARD; PRT; 2818 AA.
XX
XX xxxxxx
XX
DT XX
XX
XX Sequence 2, Application US/08411389
XX
XX Sequence 2, Application US/08411389
XX Patent No. 5605799
XX GENERAL INFORMATION:
CC APPLICANT: White, Raymond L.
CC APPLICANT: Cawthon, Richard M.
CC APPLICANT: Li, Ying
CC TITLE OF INVENTION: SOMATIC MUTATIONS IN THE
CC TITLE OF INVENTION: NEUROFIBROMATOSIS TYPE 1 GENE IN HUMAN TUMORS
CC NUMBER OF SEQUENCES: 21
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Venable, Baetjer, Howard & Civiletti
CC STREET: 1201 New York Avenue NW, Suite 1000
CC CITY: Washington
CC STATE: DC
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/411.389
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/047.088
CC FILING DATE: 16-APR-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Ihnen, Jeffrey L.
CC REGISTRATION NUMBER: 28,957
CC REFERENCE/DOCKET NUMBER: 19780-107116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-962-4810
CC TELEFAX: 202-962-8300
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 2818 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 2818 AA: 317012 MW; 42200472 CN;
Query Match 70.3%; Score 45; DB 1: Length 2818;
Best Local Similarity 60.08; Pred. No. 1.31e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1526 GPPEHKFPVAD 1535
QY :|||:|||
QY 1 APPEDNEVED 10

RESULT 11
XX US-08-449-933-2 STANDARD; PRT; 2818 AA.
XX
XX xxxxxx
XX AC
XX
XX DT
XX
XX Sequence 2, Application US/08449933
XX
XX Sequence 2, Application US/08449933
XX Patent No. 5859195
XX GENERAL INFORMATION:
CC APPLICANT: Collins, Francis S.

```

CC APPLICANT: Wallace, Margaret R.  
CC APPLICANT: Marchuk, Douglas A.  
CC APPLICANT: Anderson, Lone B.  
CC APPLICANT: Gutman, David H.  
CC TITLE OF INVENTION: Neurofibromatosis Gene  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Morrison & Foerster  
CC STREET: 755 Page Mill Road  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94304-1018  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/449,933  
CC FILING DATE: 25-MAY-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Kanski, Antoinette F.  
CC REGISTRATION NUMBER: 34,202  
CC REFERENCE/DOCKET NUMBER: 20344-20553.10  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 813-5600  
CC TELEFAX: (415) 494-0792  
CC TELEX: 706141 MRSNFOERS SFO  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2818 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA to mRNA  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC ORIGINAL SOURCE:  
CC ORGANISM: Homo sapiens  
CC POSITION IN GENOME:  
CC CHROMOSOME/SEGMENT: 17q11.2  
CC FEATURE:  
CC NAME/KEY: Cleavage-site  
CC LOCATION: group(583..586, 815..818, 2573..2576, 2810..2813)  
CC OTHER INFORMATION: /note= "Potential camp-dependent  
CC OTHER INFORMATION: protein kinase recognition sites"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: 2549..2556  
CC OTHER INFORMATION: /note= "Potential tyrosine  
CC OTHER INFORMATION: phosphorylation site"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: group(1264, 1276, 1358, 1377, 1389, 1390, 1391,  
CC LOCATION: 1395, 1396, 1400, 1423, 1426, 1429, 1430)  
CC OTHER INFORMATION: /note= "Invariant residues within  
CC OTHER INFORMATION: most statistically significant regions of similarity among  
CC OTHER INFORMATION: GAP family of proteins"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: group(1264..1290, 1345..1407, 1415..1430)  
CC OTHER INFORMATION: /note= "Most statistically  
CC OTHER INFORMATION: significant regions of similarity among the GAP family of  
CC OTHER INFORMATION: proteins"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: 496  
CC OTHER INFORMATION: /note= "At variance with previously  
CC OTHER INFORMATION: published sequence which shows an ATG methionine codon ra  
CC OTHER INFORMATION: than an ATA isoleucine codon"  
CC FEATURE:

CC NAME/KEY: Modified-site  
CC LOCATION: 1183  
CC OTHER INFORMATION: /note= "At variance with previously  
CC OTHER INFORMATION: published sequence. Shows an CTG leucine codon rather  
CC OTHER INFORMATION: previously published CTC"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: 1555  
CC OTHER INFORMATION: /note= "At variance with previously  
CC OTHER INFORMATION: published sequence. Lacks an extra CAT histidine condo  
CC OTHER INFORMATION: this residue"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: (2771-2772)  
CC OTHER INFORMATION: /note= "Position of an 18 amino  
CC OTHER INFORMATION: acid insertion(SEQ ID NO:10) representing an alternati  
CC OTHER INFORMATION: spliced product"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: (1370-1371)  
CC OTHER INFORMATION: /note= "Position of a 21 amino acid  
CC OTHER INFORMATION: insertion representing an alternatively spliced produc  
CC FEATURE:  
CC NAME/KEY: Domain  
CC LOCATION: 1125..1537  
CC OTHER INFORMATION: /note= "NFI catalytic domain"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: 2746..2818  
CC OTHER INFORMATION: /note= "Corresponding amino acids  
CC OTHER INFORMATION: for the PstI-HindIII fragment designated PMAL.B3A"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: 65..371  
CC OTHER INFORMATION: /note= "Corresponding amino acids  
CC OTHER INFORMATION: for the HpaI-PstI fragment designated PMAL.HF3A.P"  
CC FEATURE:  
CC NAME/KEY: Modified-site  
CC LOCATION: 65..1240  
CC OTHER INFORMATION: /note= "Corresponding amino acids  
CC OTHER INFORMATION: for the HpaI-XhoI fragment designated PMAL.HF3A.X"  
CC PUBLICATION INFORMATION:  
CC AUTHORS: Wallace, M.R. et al.  
CC TITLE: Type 1 Neurofibromatosis Gene: Correction  
CC JOURNAL: Science  
CC VOLUME: 250  
CC ISSUE: 12/21/90  
CC PAGES: 1749-  
CC DATE: 12/21-1990  
CC RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2818  
CC PUBLICATION INFORMATION:  
CC AUTHORS: Wallace, M.R. et al.  
CC TITLE: Type 1 Neurofibromatosis Gene: Identification  
CC TITLE: Of a Large Transcript in Three NFI Patients  
CC JOURNAL: Science  
CC VOLUME: 249  
CC ISSUE: 07/13/90  
CC PAGES: 181-186  
CC DATE: 07/13-1990  
CC RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2818  
CC SEQUENCE 2818 AA; 317011 MW; 42186697 CN;

Query Match 70.3%; Score 45; DB 2; Length 2818;  
Best Local Similarity 60.0%; Pred. No. 1.31e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1526 GPPHKPVAD 1535  
Qy 1 APPENDPVED 10

RESULT 12  
ID US-07-776-272-18 STANDARD; PRT; 36 AA.



CC GENERAL INFORMATION:  
CC APPLICANT: Kaper Dr., James B.  
CC APPLICANT: Levine Dr., Myron M.  
CC TITLE OF INVENTION: Vibrio cholerae 01 (CVD111) and non-01  
CC TITLE OF INVENTION: (CVD112 and CVD112RM) serogroup vaccine strains, methods  
CC TITLE OF INVENTION: of making same and products thereof  
CC NUMBER OF SEQUENCES: 16  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Spencer & Frank  
CC STREET: 1100 New York Ave. N.W. Suite 300 East  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/624,601  
CC FILING DATE: 08-APR-1996  
CC CLASSIFICATION: 424  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Schneller Dr., John W.  
CC REGISTRATION NUMBER: 26,031  
CC REFERENCE/DOCKET NUMBER: BAMCZ0019P2  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202)414-4000  
CC TELEFAX: (202)414-4040  
CC INFORMATION FOR SEQ ID NO: 15:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 395 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC ORIGINAL SOURCE:  
CC ORGANISM: Vibrio cholerae  
CC STRAIN: classical 395  
CC IMMEDIATE SOURCE:  
CC CLONE: orfu protein  
CC SEQUENCE 395 AA; 43922 MW; 793434 CN;  
  
Query Match 68.8%; Score 44; DB 2; Length 395;  
Best Local Similarity 55.6%; Pred. No. 1.64e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Db 121 SPPSDSPVD 129  
:|:|:|:  
Qy 1 APPEDNPVE 9  
  
RESULT 15  
ID US-07-931-943-5 STANDARD; PRT; 395 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
XX  
XX Sequence 5, Application US/07931943  
XX Sequence 5, Application US/07931943  
CC Patent No. 5470729  
CC GENERAL INFORMATION:  
CC APPLICANT: KAPER, James B.  
CC APPLICANT: BAUDRY-MAURELLI, Bernadette  
CC APPLICANT: FASANO, Alessio  
CC TITLE OF INVENTION: METHOD OF ISOLATING RESTRICTION FRAGMENT  
CC NUMBER OF SEQUENCES: 6

CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: BAKER & MCKENZIE  
CC STREET: 815 Connecticut Avenue, N.W.  
CC CITY: Washington  
CC STATE: DC  
CC COUNTRY: US  
CC ZIP: 20006  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/931,943  
CC FILING DATE: 19920812  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/821,072  
CC FILING DATE: 16-JAN-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/533,315  
CC FILING DATE: 05-JUN-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 06/581,406  
CC FILING DATE: 17-FEB-1984  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 06/472,276  
CC FILING DATE: 04-MAR-1983  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/363,383  
CC FILING DATE: 05-JUN-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 06/867,633  
CC FILING DATE: 27-MAY-1986  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Kile, Bradford E.  
CC REGISTRATION NUMBER: 25,223  
CC REFERENCE/DOCKET NUMBER: BAMCZ0016P6  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202) 452-7000  
CC TELEFAX: (202) 452-7074  
CC INFORMATION FOR SEQ ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 395 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 395 AA; 43974 MW; 802149 CN;  
  
Query Match 68.8%; Score 44; DB 1; Length 395;  
Best Local Similarity 55.6%; Pred. No. 1.64e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Db 121 SPPSDSPVD 129  
:|:|:|:  
Qy 1 APPEDNPVE 9

Search completed: Thu Oct 21 15:33:58 1999  
Job time : 8 secs.

\*\*\*\*\*  
[W][A][T][E][R][M][A][N]  
[P][R][O][T][E][I][N]  
[S][E][Q][U][E][N][C][E]  
[I][D]:[US09040485]  
[T][M]  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 21 15:31:53 1999; MasPar time 3.06 Seconds  
131.112 Million cell updates/sec

Tabular output not generated.

Title: >US-09-040-485-6  
Description: (1-10) from US09040485.pep  
Perfect Score: 64  
Sequence: 1 APPEDNPVED 10

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.188; Variance 27.251; scale 0.778

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Match	Query	ID	Description	Pred. No.	
1	64	100.0	757	2	I38423	aspartyl beta-hydroxy	1.62e-03
2	50	78.1	704	1	S60117	protein kinase C (EC	2.14e+00
3	47	73.4	129	2	A70970	hypothetical protein	8.79e+00
4	46	71.9	562	2	JC5182	serine C-palmitoyl	1.39e+01
5	45	70.3	86	2	S76775	hypothetical protein	2.19e+01
6	45	70.3	97	2	S11755	hypothetical protein	2.19e+01
7	45	70.3	258	2	C70885	probable dehydrogenas	2.19e+01
8	45	70.3	339	2	I78852	copB homolog - Xantho	2.19e+01
9	45	70.3	399	2	I78852	neurofibromatosis pro	2.19e+01
10	45	70.3	928	1	RBHU	retinoblastoma-associ	2.19e+01
11	45	70.3	932	2	F69552	leucyl-tRNA synthetas	2.19e+01
12	45	70.3	2818	2	B55282	neurofibromatosis-rel	2.19e+01
13	45	70.3	2820	2	JC5196	neurofibromin I - rat	2.19e+01
14	45	70.3	2825	2	I54352	neurofibromin - mouse	2.19e+01
15	44	68.8	36	2	A28578	pancreatic hormone -	3.41e+01
16	44	68.8	80	1	PCCH	pancreatic hormone pr	3.41e+01
17	44	68.8	122	2	D35392	hypothetical protein	3.41e+01
18	44	68.8	213	2	JQ0075	troponomodulin - goldf	3.41e+01
19	44	68.8	257	2	B31957	troponin T, skeletal	3.41e+01
20	44	68.8	259	2	F69311	conserved hypothetical	3.41e+01
21	44	68.8	263	2	C31957	troponin T, skeletal	3.41e+01
22	44	68.8	304	2	A48174	beta-lactanase (EC 3.	3.41e+01
23	44	68.8	376	2	S12730	actin - California se	3.41e+01

24 44 68.8 395 2 S36029 hypothetical protein 3.41e-01  
25 44 68.8 979 2 A35913 regulatory factor X - 3.41e-01  
26 43 67.2 176 2 S06638 calpastatin - bovine 5.29e-01  
27 43 67.2 263 2 C65044 hypothetical protein 5.29e-01  
28 43 67.2 307 2 A36885 bo-type ubiquinol oxi 5.29e-01  
29 43 67.2 336 2 T01306 hypothetical protein 5.29e-01  
30 43 67.2 491 2 A49993 glycolipptide N-tetra 5.29e-01  
31 43 67.2 506 1 FOLJG5 gag polyprotein - sim 5.29e-01  
32 43 67.2 507 2 S04237 gag polyprotein - sim 5.29e-01  
33 43 67.2 966 2 S25365 gag polyprotein - yeast 5.29e-01  
34 43 67.2 976 2 S40697 processing endoprotei 5.29e-01  
35 43 67.2 1257 2 A41060 neural cell adhesion 5.29e-01  
36 43 67.2 1259 2 S36126 neural cell adhesion 5.29e-01  
37 43 67.2 1260 2 S05479 neural cell adhesion 5.29e-01  
38 42 65.6 313 2 S26838 nodulation protein no 8.14e-01  
39 42 65.6 506 1 FOLJG3 gag polyprotein - sim 8.14e-01  
40 42 65.6 547 2 S15028 chromatin-binding pro 8.14e-01  
41 42 65.6 551 2 T00932 hypothetical protein 8.14e-01  
42 42 65.6 670 2 G71251 probable DNA helicase 8.14e-01  
43 42 65.6 837 2 D71027 hypothetical protein 8.14e-01  
44 42 65.6 909 2 S32538 cGMP-gated cation cha 8.14e-01  
45 42 65.6 1836 2 J50648 sodium channel alpha 8.14e-01

ALIGNMENTS

RESULT 1  
ENTRY I38423 #type complete  
TITLE aspartyl beta-hydroxylase - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 10-Jul-1998

ACCESSIONS I38423  
REFERENCE I38423  
#authors Koriath, F.; Gieffers, C.; Frey, J.  
#journal Gene (1994) 150:395-399  
#title Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase.  
#cross-references MUID:95121937  
#accession I38423

##status preliminary; translated from GB/EMBL/DBDJ  
##molecule\_type mRNA  
##residues 1-757 #label RES  
##cross-references EMBL:U03109; NID:9458031; PID:9458032  
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology

FEATURE 54-75  
SUMMARY #domain transmembrane #status predicted #label TRM #length 757 #molecular-weight 85498 #checksum 2143

Query Match 100.0%; Score 64; DB 2; Length 757;  
Best Local Similarity 100.0%; Pred. No. 1.62e-03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 281 APPEDNPVED 290  
QY 1 APPEDNPVED 10  
|||||||

RESULT 2  
ENTRY S60117 #type complete  
TITLE protein kinase C (EC 2.7.1.-) TPA-1A - Caenorhabditis elegans  
CONTAINS protein kinase C TPA-1B  
ORGANISM #formal\_name Caenorhabditis elegans  
DATE 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 05-Sep-1997

ACCESSIONS S60117  
REFERENCE S60117  
#authors Sano, T.; Tabuse, Y.; Nishiwaki, K.; Miwa, J.  
#journal J. Mol. Biol. (1995) 251:477-485  
#title The tpa-1 gene of Caenorhabditis elegans encodes two proteins similar to ca(2+)-independent protein kinase Cs: evidence by complete genomic and complementary DNA sequences of the



```

tpa-1 gene.
#cross-references MUID:95387388
#accession S60117
##molecule_type DNA
##residues 1-704 ##label SAN
##cross-references EMBL:D49525; NID:g1217583; PID:d1009081; PID:g1217584
GENETICS
#gene tpa-1
#map_position IV
#introns 9/2: 45/1; 93/3; 128/1; 153/1; 226/3; 270/1; 446/3; 498/2; 655/3
FUNCTION
#description catalyzes the formation of peptidyl-serine-phosphate or
peptidyl-threonine-phosphate using ATP
#note activity is calcium-independent, phospholipid-dependent, and
activated by diacylglycerol and by tumor-promoting phorbol
esters
CLASSIFICATION
#superfamily protein kinase C delta; protein kinase C
zinc-binding repeat homology; protein kinase homology
alternative splicing; ATP; duplication; phorbol ester
binding; phospholipid binding; phosphotransferase;
serine/threonine-specific protein kinase; zinc
FEATURE
138-704 #product protein kinase C TPA-1B #status predicted
#label MAT2\
#region pseudophosphorylation motif\
#domain protein kinase C zinc-binding repeat homology
#label K21\
#domain protein kinase C zinc-binding repeat homology
#label K22\
#domain protein kinase C zinc-binding repeat homology
#domain protein kinase homology #label KIN\
#region protein kinase ATP-binding motif\
#binding_site zinc (His, Cys, Cys, Cys) #status
predicted\
#binding_site zinc (Cys, Cys, His, Cys) #status
predicted\
#binding_site zinc (His, Cys, Cys, Cys) #status
predicted\
#binding_site zinc (Cys, Glu, Asp, Lys #status predicted
#active_site Lys, Glu, Asp, Lys #status predicted
SUMMARY #length 704 #molecular-weight 80298 #checksum 7955
Query Match 78.1%; Score 50; DB 1; Length 704;
Best Local Similarity 70.0%; Pred. No. 2.14e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 127 APDDDPVED 136
QY 1 APPEDNPVED 10
RESULT 3
ENTRY #type complete
TITLE A70970 #type complete
hypoetical protein Rv3354 - Mycobacterium tuberculosis
(strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS A70970
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrrell, B.G.
Nature (1998) 393:537-544
Deciphering the Biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession A70970
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-129 ##label COL
##cross-references GB:AL009198; GB:AL123456; NID:g3242262; PID:e1202271;
PID:g2661632
##experimental_source strain H37RV
GENETICS
#gene Rv3354
#summary #length 129 #molecular-weight 12988 #checksum 6975
Query Match 73.4%; Score 47; DB 2; Length 129;
Best Local Similarity 60.0%; Pred. No. 8.79e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 28 APAQANPVDD 37
QY 1 APPEDNPVED 10
RESULT 4
ENTRY #type complete
TITLE JC5182 #type complete
serine C-palmitoyltransferase (EC 2.3.1.50) Lcb2 chain -
yeast (Kluyveromyces fragilis)
ALTERNATE_NAMES 3-ketosphinganine synthetase; SPT
ORGANISM #formal_name Kluyveromyces fragilis
DATE 16-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
17-Mar-1999
ACCESSIONS JC5182
REFERENCE JC5180
#authors Nagiec, M.M.; Lester, R.L.; Dickson, R.C.
#journal Gene (1996) 177:237-241
#title Sphingolipid synthesis: Identification and characterization
of mammalian cDNAs encoding the Lcb2 subunit of serine
palmitoyltransferase.
#cross-references MUID:97080528
#accession JC5182
##status preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-562 ##label NAG
##cross-references GB:U15646; NID:g1001948; PID:g1001949
COMMENT This enzyme catalyzes the irreversible reaction of condensation of
serine and palmitoyl-CoA to yield 3-ketosphinganine.
CLASSIFICATION #superfamily serine C-palmitoyltransferase chain LCB2
KEYWORDS acyltransferase
FEATURE 319-374
SUMMARY #length 562 #molecular-weight 63004 #checksum 5581
Query Match 71.9%; Score 46; DB 2; Length 562;
Best Local Similarity 77.8%; Pred. No. 1.39e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 13 PPEDIPLED 21
QY 2 PPEDNPVED 10
RESULT 5
ENTRY #type complete
TITLE S76775 #type complete
hypoetical protein - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S76775
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;

```

```

Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpou,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
DNA Res. (1996) 3:109-136
#journal
#title
Sequence analysis of the genome of the unicellular
cyanobacterium Synechocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S76775
#status preliminary
#molecule_type DNA
#residues 1-86 #label KAN
##cross-references EMBL:D90916; GB:AB001339; NID:g1653715; PID:d1019420;
PID:g1653776
#note the nucleotide sequence was submitted to the EMBL Data
Library June 1996
SUMMARY #length 86 #molecular-weight 9909 #checksum 7709
Query Match 70.3%; Score 45; DB 2; Length 86;
Best Local Similarity 62.5%; Pred. No. 2.19e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 49 PDDTPVEE 56
I:| | | |
Qy 3 PEDNPVED 10

RESULT 6
ENTRY S11755 #type fragment
TITLE hypothetical protein - Plasmodium vivax (fragment)
ORGANISM #formal_name Plasmodium vivax
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
09-Sep-1997
ACCESSIONS S11755
REFERENCE S11755
#authors Campbell, J.R.; Franke, E.D.
#submission submitted to the EMBL Data Library, April 1989
#description A repetitive element found in Plasmodium vivax DNA.
#accession S11755
#molecule_type DNA
#residues 1-97 #label CAM
##cross-references EMBL:X15129; NID:g10086; PID:g10087
SUMMARY #length 97 #checksum 7475
Query Match 70.3%; Score 45; DB 2; Length 97;
Best Local Similarity 71.4%; Pred. No. 2.19e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 13 PPEENPI 19
I:| | | |
Qy 2 PPEDNPV 8

RESULT 7
ENTRY C70885 #type complete
TITLE probable dehydrogenase - Mycobacterium tuberculosis (strain
H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
12-Feb-1999
ACCESSIONS C70885
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skelton, S.; Squares, S.; Squares, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.

```

```

#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession C70885
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-258 #label COL
##cross-references GB:AL008883; GB:AL123456; NID:g3261490; PID:e1172949;
PID:g2612802
##experimental_source strain H37Rv
GENETICS
#gene RV2857c
CLASSIFICATION #superfamily short-chain alcohol dehydrogenase homology
FEATURE
11-188 #domain short-chain alcohol dehydrogenase homology
#label SADH
SUMMARY #length 258 #molecular-weight 26804 #checksum 1750
Query Match 70.3%; Score 45; DB 2; Length 258;
Best Local Similarity 60.0%; Pred. No. 2.19e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 93 SPEDNLIEN 102
I:| | | | | |
Qy 1 APPEDNPVED 10

RESULT 8
ENTRY B36868 #type complete
TITLE copB homolog - Xanthomonas campestris
ALTERNATE_NAMES hypothetical protein 2
ORGANISM #formal_name Xanthomonas campestris
DATE 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change
09-Sep-1997
ACCESSIONS B36868
REFERENCE A36868
#authors Lee, Y.A.; Hendson, M.; Panopoulos, N.J.; Schroth, M.N.
#journal J. Bacteriol. (1994) 176:173-188
#title Molecular cloning, chromosomal mapping, and sequence analysis
of copper resistance genes from Xanthomonas campestris pv.
juglandis: homology with small blue copper proteins and
multicopper oxidase.
#cross-references MUID:94110224
#accession B36868
#status preliminary
#molecule_type DNA
#residues 1-339 #label LEE
##cross-references GB:L19222; NID:g349160; PID:g461142
SUMMARY #length 339 #molecular-weight 37041 #checksum 1404
Query Match 70.3%; Score 45; DB 2; Length 339;
Best Local Similarity 60.0%; Pred. No. 2.19e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 121 PPPSDHPVHD 130
I:| | | | |
Qy 1 APPEDNPVED 10

RESULT 9
ENTRY I78852 #type fragment
TITLE neurofibromatosis protein type 1 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
29-Aug-1997
ACCESSIONS I78852
REFERENCE I58356
#authors Nishi, T.; Lee, P.S.; Oka, K.; Levin, V.A.; Tanase, S.;
Morino, Y.; Saya, H.
#journal Oncogene (1991) 6:1555-1559
#title Differential expression of two types of the neurofibromatosis

```

type 1 (NF1) gene transcripts related to neuronal differentiation.

#cross-references MUID:92019823

#accession I78852

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type mRNA

##residues 1-399 #label RES

##cross-references GB:M60915; NID:g189159; PID:g189161

#### GENETICS

#gene GDB:NFI

##cross-references GDB:I20231; OMIM:162200

#map\_position 17q11.2-17q11.2

CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology

FEATURE #domain ras-specific GAP catalytic domain homology

68-303

#label GAP

SUMMARY #length 399 #checksum 4004

Query Match 70.3%; Score 45; DB 2; Length 399;

Best Local Similarity 60.0%; Pred.No. 2.19e+01;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 380 GPEKPKVAD 389

QY 1 APPEDNPVED 10

|||||

RESULT 10

ENTRY RBHU #type complete

TITLE retinoblastoma-associated protein - human

ALTERNATE\_NAMES retinoblastoma susceptibility protein

ORGANISM #formal\_name Homo sapiens #common\_name man

DATE 30-Jun-1987 #sequence\_revision 30-Jun-1990 #text\_change

ACCESSIONS JS0276; A03152; A91613; A39947; A44987; I54364; I58362;

I78863; I78866; I78872; I78873; A35590

REFERENCE JS0276

#authors Lee, W.H.; Shew, J.Y.; Hong, F.D.; Sery, T.W.; Donoso, L.A.;

Young, L.J.; Bookstein, R.; Lee, E.Y.H.P.

#journal Nature (1987) 329:642-645

#title The retinoblastoma susceptibility gene encodes a nuclear

phosphoprotein associated with DNA binding activity.

#cross-references MUID:88014238

#accession JS0276

##molecule\_type mRNA

##residues 1-928 #label LE1

##cross-references GB:M28419; NID:g190962; PID:g190963

##note this sequence has two possible initiation sites, 1-Met

and 113-Met

REFERENCE A03152

#authors Lee, W.H.; Bookstein, R.; Hong, F.; Young, L.J.; Shew, J.Y.;

Lee, E.Y.H.P.

#journal Science (1987) 235:1394-1399

#title Human retinoblastoma susceptibility gene: cloning,

identification, and sequence.

#cross-references MUID:87149066

#accession A03152

##molecule\_type mRNA

##residues 113-116, 'LLSYRXTY', 125-332, 'R', 334-367, 'I', 369-928

#label LE2

##cross-references GB:M15400; NID:g190958; PID:g190959

REFERENCE A91613

#authors McGee, T.L.; Vandell, D.W.; Dryja, T.P.

#journal Gene (1989) 80:119-128

#title Structure and partial genomic sequence of the human

retinoblastoma susceptibility gene.

#cross-references MUID:90006771

#accession A91613

##molecule\_type DNA

##residues 1-928 #label MCG

##cross-references GB:M27845; GB:L11910; NID:g292420; PID:g292421

##note the authors translated the codon GAA for residue 559 as

Gly

#### REFERENCE

#authors

Friend, S.H.; Horowitz, J.M.; Gerber, M.R.; Wang, X.F.;

Bogenmann, E.; Li, F.P.; Weinberg, R.A.

#journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:9059-9063

#title Deletions of a DNA sequence in retinoblastomas and

mesenchymal tumors: organization of the sequence and its

encoded protein.

#cross-references MUID:88097427

#accession A39947

##molecule\_type mRNA

##residues 1-928 #label FRI

##cross-references GB:M33647; GB:J02994; NID:g190945; PID:g190946

REFERENCE A44987

#authors

T'Ang, A.; Wu, K.J.; Hashimoto, T.; Liu, W.Y.; Takahashi, R.;

Shi, X.H.; Mihara, K.; Zhang, F.H.; Chen, Y.Y.; Du, C.;

Qian, J.; Lin, Y.G.; Murphree, A.L.; Qiu, W.R.; Thompson,

T.; Benedict, W.F.; Fung, Y.K.T.

#journal Oncogene (1989) 4:401-407

#title Genomic organization of the human retinoblastoma gene.

#cross-references MUID:89239464

#accession A44987

##molecule\_type DNA

##residues 1-46 #label TAA

##cross-references EMBL:X16439; NID:g35894; PID:g35895

REFERENCE I54364

#authors

Lohmann, D.R.; Brandt, B.; Hopping, W.; Passarge, E.;

Horsthemke, B.

#journal Hum. Mol. Genet. (1994) 3:2187-2193

#title Spectrum of small length germline mutations in the RB1 gene.

#cross-references MUID:95187159

#accession I54364

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 128-133 #label LOH

##cross-references GB:L49209; NID:g1088286; PID:g1088287

REFERENCE I58362

#authors

Hogg, A.; Onadim, Z.; Baird, P.N.; Cowell, J.K.

#journal Oncogene (1992) 7:1445-1451

#title Detection of heterozygous mutations in the RB1 gene in

retinoblastoma patients using single-strand conformation

polymorphism analysis and polymerase chain reaction

sequencing

#cross-references MUID:92319557

#accession I58362

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 1-45 #label RE2

##cross-references GB:L41889; NID:g793948; PID:g793949

#accession I78863

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 377-394 #label RE5

##cross-references GB:L41900; NID:g793969; PID:g793972

#accession I78866

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 655-671 #label RE4

##cross-references GB:L41907; NID:g801729; PID:g801730

#accession I78872

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type DNA

##residues 906-928 #label RE3

##cross-references GB:L41914; NID:g794010; PID:g794011

REFERENCE A35590

#authors

Lee, E.Y.H.P.; Bookstein, R.; Young, L.J.; Lin, C.J.;

Rosenfeld, M.G.; Lee, W.H.

#journal Proc. Natl. Acad. Sci. U.S.A. (1988) 85:6017-6021

#title Molecular mechanism of retinoblastoma gene inactivation in

```

retinoblastoma cell line Y79.
#cross-references MUID:88320373
#contents
#annotation
REFERENCE A38988
#authors Lees, J.A.; Buchkovich, K.J.; Marshak, D.R.; Anderson, C.W.; Harlow, E.
#journal EMBO J. (1991) 10:4279-4290
#title The retinoblastoma protein is phosphorylated on multiple sites by human cdc2.
#cross-references MUID:92097548
#contents
#annotation: phosphorylation sites
#gene GDB:RB1
##cross-references GDB:118734; OMIM:180200
#map_position 13q14.3-13q14.3
#intons 46/2; 88/3; 127/2; 167/2; 180/2; 203/1; 240/1; 287/3; 313/3; 350/2; 376/2; 405/3; 444/3; 463/3; 474/2; 500/1; 565/3; 605/2; 634/1; 702/3; 737/3; 830/2; 840/3; 888/2; 905/1
#note mutations in the germline gene predispose to hereditary retinoblastoma; somatic mutations have been found in a variety of human cancers including retinoblastoma and osteosarcoma
CLASSIFICATION #superfamily retinoblastoma-associated protein
KEYWORDS cell cycle control; DNA binding; leucine zipper; osteosarcoma; phosphoprotein; retinoblastoma; transcription regulation; tumor suppressor
FEATURE
10-18 #region alanine-rich
20-29 #region proline-rich
662-683 #region leucine zipper motif
249, 807, 811 #binding_site phosphate (Ser) (covalent) (by cdc2 kinase) #status predicted
252, 373 #binding_site phosphate (Thr) (covalent) (by cdc2 kinase) #status predicted
SUMMARY #length 928 #molecular-weight 106158 #checksum 3504
Query Match 70.3%; Score 45; DB 1; Length 928;
Best Local Similarity 50.0%; Pred. No. 2.19e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 27 PPEDPQD 36
QY 1 APPEDNPVED 10
:||||:|
RESULT 11
ENTRY #type complete
TITLE leucyl-tRNA synthetase (leus) homolog - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Aug-1998
ACCESSION F69552
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Keriavage, A.R.; Graham, D.E.; Kyriades, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
Nature (1997) 390:364-370
#journal The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
#title sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession F69552
#status preliminary
#molecule_type DNA
##residues 1-932 #label KLE
##cross-references GB:AE001108; GB:AE000782; NID:g2689431; PID:g2650668; TIGR:AF2421
CLASSIFICATION #superfamily valine--tRNA ligase
SUMMARY #length 932 #molecular-weight 108627 #checksum 6668
Query Match 70.3%; Score 45; DB 2; Length 932;
Best Local Similarity 66.7%; Pred. No. 2.19e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 180 PHDONPVED 188
QY 2 PPEDNPVED 10
|::|:||||
RESULT 12
ENTRY #type complete
TITLE neurofibromatosis-related protein NF1 - human
ALTERNATE_NAMES Gpase activating protein homolog NF1; neurofibromin
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 20-Mar-1998
ACCESSION B55282; A55282; A35879; A35605; A35910; A35222; A36297; I58356
REFERENCE A55282
#authors Marchuk, D.A.; Saulino, A.M.; Tavakkol, R.; Swaroop, M.; Wallace, M.R.; Andersen, L.B.; Mitchell, A.L.; Gutmann, D.H.; Boguski, M.; Collins, F.S.
#journal Genomics (1991) 11:931-940
#title cDNA cloning of the type 1 neurofibromatosis gene: complete sequence of the NF1 gene product.
#cross-references MUID:92147138
#accession B55282
#status not compared with conceptual translation
#molecule_type mRNA
##residues 1-2818 #label MAR
##cross-references GB:M82814; NID:g189164; PID:g189165
#note sequence extracted from NCBI backbone (NCBIP:80176)
#accession A55282
#status preliminary
#molecule_type mRNA
##residues 1-334 #label MA2
##note sequence extracted from NCBIP backbone (NCBIN:80169, NCBIP:80172)
REFERENCE A35879
#authors Xu, G.; O'Connell, P.; Viskochil, D.; Cawthon, R.; Robertson, M.; Culver, M.; Dunn, D.; Stevens, J.; Gesteland, R.; White, R.; Weiss, R.
#journal Cell (1990) 62:599-608
#title The neurofibromatosis type 1 gene encodes a protein related to GAP.
#cross-references MUID:90335969
#accession A35879
#status preliminary
#molecule_type mRNA
##residues 335-495, '1', 497-1555, 'H', 1556-2818 #label XUA
##cross-references GB:M38106; GB:M57449; NID:g189169; PID:g189170
REFERENCE A35605
#authors Cawthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.; Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.; O'Connell, P.; White, R.
#journal Cell (1990) 62:193-201
#title A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations.
#cross-references MUID:90304909
#accession A35605
#status preliminary
#molecule_type mRNA
##residues 1585-2687 #label CAW
##cross-references EMBL:M38107; EMBL:M57449

```

```

REFERENCE A35910
#authors Cawthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
O'Connell, P.; White, R.
#journal Cell (1990) 62:608b
#accession A35910
#status preliminary; nucleic acid sequence not shown; not
compared with conceptual translation
##molecule_type mRNA 2688-2818 #label CA2
##residues A35222
#authors Wallace, M.R.; Marchuk, D.A.; Andersen, L.B.; Letcher, R.;
Odeh, H.M.; Saulino, A.M.; Fountain, J.W.; Brereton, A.;
Nicholson, J.; Mitchell, A.L.; Brownstein, B.H.; Collins,
P.S.
#journal Science (1990) 249:181-186
#title Type 1 neurofibromatosis gene: identification of a large
transcript disrupted in three NF1 patients.
#cross-references MUID:90319792
#accession A35222
#status preliminary
##molecule_type mRNA
##residues 2209-2818 #label WAL
##cross-references GB:M60496; NID:g189157; PID:g189158; GB:M49193
REFERENCE A36297
#authors Martin, G.A.; Viskochil, D.; Bollag, G.; McCabe, P.C.;
Crosier, W.J.; Haubruck, H.; Conroy, L.; Clark, R.;
O'Connell, P.; Cawthon, R.M.; Innis, M.A.; McCormick, F.
#journal Cell (1990) 63:843-849
#title The GAP-related domain of the neurofibromatosis type 1 gene
product interacts with ras p21.
#cross-references MUID:91029515
#accession A36297
##molecule_type mRNA
##residues 1096-1569, 'TPPPEP' #label MA3
##cross-references GB:M61213; NID:g189162; PID:g189163
##note this clone includes an epitope tag at the 3' end
encoding the sequence TPPPEP, not part of dystrophin
but recognized by the monoclonal antibody KT3.
REFERENCE I58356
#authors Nishi, T.; Lee, P.S.; Oka, K.; Levin, V.A.; Tanase, S.;
Morino, Y.; Sava, H.
#journal Oncogene (1991) 6:1555-1559
#title Differential expression of two types of the neurofibromatosis
type 1 (NF1) gene transcripts related to neuronal
differentiation.
#cross-references MUID:92019823
#accession I58356
#status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1168-1545 #label RES
##cross-references GB:M60915; NID:g189159; PID:g189160
GENETICS
#gene GDB:NFI
##cross-references GDB:120231; OMIM:162200
#map_position 17q11.2-17q11.2
#introns 1370/3
#note the list of introns is incomplete
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
KEYWORDS alternative splicing; tumor suppressor
FEATURE
1235-1449 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY #length 2818 #molecular-weight 317030 #checksum 2858
Query Match 70.3%; Score 45; DB 2; Length 2818;
Best Local Similarity 60.0%; Pred. No. 2.19e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1526 GPPCHKPVAD 1535
:|||||
QY 1 APPEDNPVED 10
:|||||

REFERENCE A35910
#authors Cawthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
O'Connell, P.; White, R.
#journal Cell (1990) 62:608b
#accession A35910
#status preliminary; nucleic acid sequence not shown; not
compared with conceptual translation
##molecule_type mRNA 2688-2818 #label CA2
##residues A35222
#authors Wallace, M.R.; Marchuk, D.A.; Andersen, L.B.; Letcher, R.;
Odeh, H.M.; Saulino, A.M.; Fountain, J.W.; Brereton, A.;
Nicholson, J.; Mitchell, A.L.; Brownstein, B.H.; Collins,
P.S.
#journal Science (1990) 249:181-186
#title Type 1 neurofibromatosis gene: identification of a large
transcript disrupted in three NF1 patients.
#cross-references MUID:90319792
#accession A35222
#status preliminary
##molecule_type mRNA
##residues 2209-2818 #label WAL
##cross-references GB:M60496; NID:g189157; PID:g189158; GB:M49193
REFERENCE A36297
#authors Martin, G.A.; Viskochil, D.; Bollag, G.; McCabe, P.C.;
Crosier, W.J.; Haubruck, H.; Conroy, L.; Clark, R.;
O'Connell, P.; Cawthon, R.M.; Innis, M.A.; McCormick, F.
#journal Cell (1990) 63:843-849
#title The GAP-related domain of the neurofibromatosis type 1 gene
product interacts with ras p21.
#cross-references MUID:91029515
#accession A36297
##molecule_type mRNA
##residues 1096-1569, 'TPPPEP' #label MA3
##cross-references GB:M61213; NID:g189162; PID:g189163
##note this clone includes an epitope tag at the 3' end
encoding the sequence TPPPEP, not part of dystrophin
but recognized by the monoclonal antibody KT3.
REFERENCE I58356
#authors Nishi, T.; Lee, P.S.; Oka, K.; Levin, V.A.; Tanase, S.;
Morino, Y.; Sava, H.
#journal Oncogene (1991) 6:1555-1559
#title Differential expression of two types of the neurofibromatosis
type 1 (NF1) gene transcripts related to neuronal
differentiation.
#cross-references MUID:92019823
#accession I58356
#status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1168-1545 #label RES
##cross-references GB:M60915; NID:g189159; PID:g189160
GENETICS
#gene GDB:NFI
##cross-references GDB:120231; OMIM:162200
#map_position 17q11.2-17q11.2
#introns 1370/3
#note the list of introns is incomplete
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
KEYWORDS alternative splicing; tumor suppressor
FEATURE
1235-1449 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY #length 2818 #molecular-weight 317030 #checksum 2858
Query Match 70.3%; Score 45; DB 2; Length 2818;
Best Local Similarity 60.0%; Pred. No. 2.19e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1526 GPPCHKPVAD 1535
:|||||
QY 1 APPEDNPVED 10
:|||||

RESULT 13
ENTRY JC5196 #type complete
TITLE neurofibromin 1 - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
10-Sep-1997
ACCESSIONS JC5196
REFERENCE JC5196
#authors Suzuki, H.; Takahashi, K.; Yasumoto, K.; Fuse, N.; Shibahara,
S.
#journal J. Biochem. (1996) 120:1048-1054
#title Differential tissue-specific expression of neurofibromin
isoform mRNAs in rat.
#accession JC5196
#status preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-2820 #label SUZ
##cross-references DDBJ:D45201; NID:g1841313; PID:d1008732; PID:g1841314
COMMENT This protein contains a GTPase-activating protein-related domain
which is responsible for the stimulatory effect of neurofibromin
on the tyrosinase promoter activity.
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
1177-1436 #domain GTPase-activating protein related #status
predicted #label GRD\
1237-1451 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY #length 2820 #molecular-weight 317080 #checksum 6628
Query Match 70.3%; Score 45; DB 2; Length 2820;
Best Local Similarity 60.0%; Pred. No. 2.19e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1528 GPPCHKPVAD 1537
:|||||
QY 1 APPEDNPVED 10
:|||||

RESULT 14
ENTRY I54352 #type fragment
TITLE neurofibromin - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
25-Apr-1997
ACCESSIONS I54352
REFERENCE I54352
#authors Bernards, A.; Snijders, A.J.; Hannigan, G.E.; Murthy, A.E.;
Gusella, J.F.
#journal Hum. Mol. Genet. (1993) 2:645-650
#title Mouse neurofibromatosis type 1 cDNA sequence reveals high
degree of conservation of both coding and non-coding mRNA
segments.
#cross-references MUID:93357730
#accession I54352
#status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-2825 #label RES
##cross-references GB:L10370; NID:g309452; PID:g309453
GENETICS
#gene NFI
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
1221-1456 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY #length 2825 #checksum 6076
Query Match 70.3%; Score 45; DB 2; Length 2825;
Best Local Similarity 60.0%; Pred. No. 2.19e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1533 GPPCHKPVAD 1542
:|||||
QY 1 APPEDNPVED 10
:|||||

```

QY 1 APPEDNPVED 10

```

RESULT 15
ENTRY A28578 #type complete
TITLE pancreatic hormone - ostrich
ALTERNATE_NAMES pancreatic polypeptide
ORGANISM #formal_name Struthio camelus #common_name ostrich
DATE 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change
12-Apr-1995
ACCESSIONS A28578
REFERENCE A28578
#authors Litthauer, D.; Oelofsen, W.
#journal Int. J. Pept. Protein Res. (1987) 29:739-745
#title Purification and primary structure of ostrich pancreatic
polypeptide.
#cross-references MUID:87307111
#accession A28578
#molecule_type protein
#residues 1-36 ##label LIT
##note the sequence of residues 22-23 was reported as Asn-Asp
in Fig. 7 and as Asp-Asn in Fig. 8
CLASSIFICATION #superfamily pancreatic hormone
SUMMARY #length 36 #molecular-weight 4209 #checksum 2998

Query Match 68.8%; Score 44; DB 2; Length 36;
Best Local Similarity 66.7%; Pred. No. 3.41e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 8 PGDDAPVED 16
|:|
QY 2 PPEDNPVED 10

```

Search completed: Thu Oct 21 15:32:08 1999  
Job time : 15 secs.

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
\*\*\*\*\*  
Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
Run on: Thu Oct 21 15:32:26 1999; MasPar time 2.28 Seconds  
Tabular output not generated. 123.898 Million cell updates/sec  
\*\*\*\*\*  
Title: >US-09-040-485-6  
Description: (1-10) from US09040485.p  
Perfect Score: 64  
Sequence: 1 APPEDNPVED 10  
Scoring table: PAM 150  
Gap 15  
Searched: 77977 seqs, 28268293 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot37  
1:swissprot  
Statistics: Mean 21.883; Variance 24.731; scale 0.885  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	64	100.0	1	ASPH_HUMAN	ASPARTYL/ASPARAGINYL B
2	50	78.1	1	KPC1_CAEEL	PROTEIN KINASE C-LIKE
3	46	71.9	1	LCB2_KLULA	SERINE PALMITOYLTRANSF
4	45	70.3	1	TRBP_MOUSE	PROTAMINE-1 RNA BINDIN
5	45	70.3	1	RB_HUMAN	RETINOBLASTOMA-ASSOCIA
6	45	70.3	1	SYL_ARCFU	LEUCYL-TRNA SYNTHETASE
7	45	70.3	1	NFL_HUMAN	NEUROFIBROMIN (NEUROFI
8	45	70.3	1	NFL_MOUSE	NEUROFIBROMIN (NEUROFI
9	44	68.8	1	PAHO_STRCA	PANCREATIC HORMONE (PA
10	44	68.8	1	PAHO_LARAR	PANCREATIC HORMONE (PA
11	44	68.8	1	PAHO_CHICK	PANCREATIC HORMONE PRE
12	44	68.8	1	CLP1_HELAM	LARVAL CUTICLE PROTEIN
13	44	68.8	1	YLC3_YEREN	HYPOTHETICAL 13.7 KD P
14	44	68.8	1	NEUM_CARAU	NEUROMODULIN (AXONAL M
15	44	68.8	1	TRT3_CHICK	TROPONIN T, FAST SKELE
16	44	68.8	1	ACTM_APLCA	ACTIN, MUSCLE.
17	44	68.8	1	YACE_VLEBCH	HYPOTHETICAL 44.0 KD P
18	44	68.8	1	YAU2_SCHPO	HYPOTHETICAL 52.9 KD P
19	44	68.8	1	DPOG_CHICK	DNA POLYMERASE GAMMA (
20	44	68.8	1	REF1_HUMAN	MHC CLASS II REGULATOR
21	43	67.2	1	OOX2_ACEAC	UBIQUINOL OXIDASE POLY
22	43	67.2	1	NWT_CRYNE	GLYCYLPEPTIDE N-TETRA
23	43	67.2	1	GAG_SIVNK	GAG POLYPROTEIN [CONTA

Result No.	Score	Query Match	ID	Description	Pred. No.
24	43	67.2	1	GAG_SIVS4	GAG POLYPROTEIN [CONTA
25	43	67.2	1	GAG_SIVSP	GAG POLYPROTEIN [CONTA
26	43	67.2	1	ICAL_BOVIN	CALPAIN INHIBITOR (CAL
27	43	67.2	1	SSN6_YEAST	GLUCOSE REPRESSION MED
28	43	67.2	1	XPR6_YARLI	DIBASIC PROCESSING END
29	43	67.2	1	CAML_HUMAN	NEURAL CELL ADHESION M
30	43	67.2	1	CAML_RAT	NEURAL CELL ADHESION M
31	43	67.2	1	CAML_MOUSE	NEURAL CELL ADHESION M
32	42	65.6	1	TRY6_ANOGA	TRYPSIN 6 PRECURSOR (E
33	42	65.6	1	NOD3_RHIME	MODULATION PROTEIN D I
34	42	65.6	1	GBB3_CHICK	G2/MITOTIC-SPECIFIC CY
35	42	65.6	1	YNQ6_YEAST	HYPOTHETICAL 49.7 KD P
36	42	65.6	1	GLNA_ARCFU	GLUTAMINE SYNTHETASE (
37	42	65.6	1	GAG_SIVM1	GAG POLYPROTEIN [CONTA
38	42	65.6	1	RCC_DROME	REGULATOR OF CHROMOSOM
39	42	65.6	1	CDCH_HALSA	CDCH PROTEIN.
40	42	65.6	1	SCH9_YEAST	CAMP-DEPENDENT PROTEIN
41	42	65.6	1	TOPI_ECOLI	DNA TOPOISOMERASE I (E
42	42	65.6	1	CNG4_HUMAN	CYCLOC-NUCLEOTIDE-GATE
43	42	65.6	1	YSX7_CAEEL	HYPOTHETICAL 113.1 KD
44	42	65.6	1	CIN4_RAT	SODIUM CHANNEL PROTEIN
45	42	65.6	1	POLG_LANVT	GENOME POLYPROTEIN [CO

ALIGNMENTS

RESULT	1	ASPH_HUMAN	STANDARD;	PRT;	757 AA.
ID	ASPH_HUMAN	STANDARD;	PRT;	757 AA.	
AC	Q12797;				
DT	01-NOV-1997 (REL. 35, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE)				
DE	HYDROXYLASE)				
DE	DIOXYGENASE)				
GN	ASPH.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;				
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.				
OC	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 95121937.				
RA	KORIOTH F., GIEFFERS C., FREY J.;				
RT	"Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase."				
RL	GENE 150:395-399(1994).				
CC	!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.				
CC	!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) - PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).				
CC	!- COFACTOR: IRON.				
CC	!- SUBUNIT: MONOMER (BY SIMILARITY).				
CC	!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.				
CC	!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.				
CC	!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
CC	EMBL; U03109; G458032; -				
DR	MIM; 600582; -				
DR	OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR; ENDOPLASMIC RETICULUM.				
KW	DOMAIN 1 54				
FT	CYTOPLASMIC (POTENTIAL).				

FT TRANSMEM 55 75 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 76 757 LUMENAL (POTENTIAL).  
 FT DOMAIN 13 20 POLY-SER.  
 FT DOMAIN 323 332 POLY-LYS.  
 FT CARBOHYD 452 452 POTENTIAL.  
 FT CARBOHYD 705 705 POTENTIAL.  
 SQ SEQUENCE 757 AA; 85498 MW; AE6AFC24 CRC32;

Query Match 100.0%; Score 64; DB 1; Length 757;  
 Best Local Similarity 100.0%; Pred. No. 2.44e-04;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 281 APPEDNPVED 290  
 QY 1 APPEDNPVED 10  
 |||||

RESULT 2  
 ID KPC1.CAEEL STANDARD; PRT; 704 AA.  
 AC P34722;  
 DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE PROTEIN KINASE C-LIKE 1 (EC 2.7.1.1-) (PKC).  
 GN TPA-1.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEIA; RHABDITIA; RHABDITIDA;  
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2.  
 RX MEDLINE; 95387388.  
 RA SANO T., TABUSE Y., NISHIWAKI K., MIWA J.;  
 RT "The tpa-1 gene of *Caenorhabditis elegans* encodes two proteins  
 similar to Ca(2+)-independent protein kinase Cs: evidence by complete  
 genomic and complementary DNA sequences of the tpa-1 gene.";  
 RL J. MOL. BIOL. 251:477-485(1995).  
 RN [2]  
 RP SEQUENCE OF 148-704 FROM N.A.  
 RA MEDLINE; 89186920.  
 RX TABUSE Y., NISHIWAKI K., MIWA J.;  
 RT "Mutations in a protein kinase C homolog confer phorbol ester  
 resistance on *Caenorhabditis elegans*.";  
 RL SCIENCE 243:1713-1716(1989).  
 RN [3]  
 RP REVISIONS TO REF.2.  
 RA MIWA J.;  
 RL SUBMITTED (MAR-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN  
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS  
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.  
 CC -!- ALTERNATIVE PRODUCTS: TWO FORMS (TPA-1A AND TPA-1B) ARE PRODUCED  
 CC BY ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS  
 CC THAT OF TPA-1A.  
 CC -!- SIMILARITY: CONTAINS TWO COPIES OF THE ZINC-DEPENDENT PHORBOL-  
 CC ESTER AND DAG BINDING DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D49525; G1217584; -  
 CC DR D49525; G1217585; -  
 CC EMBL; D14815; G303529; -  
 CC DR D14815; G303529; -  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 CC DR PROSITE; PS00479; DAG\_PE\_BINDING\_DOMAIN; 2.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.

DR PFAM; PF00069; pkinese; 1.  
 DR PFAM; PF00130; DAG\_PE-bind; 2.  
 DR PFAM; PF00433; pkinese\_C; 1.  
 DR HSP; P28867; IPTR.  
 KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING; ZINC;  
 KW PHORBOL-ESTER BINDING; REPEAT; ALTERNATIVE SPLICING; PHOSPHORYLATION.  
 FT DOMAIN 166 215 PHORBOL-ESTER AND DAG BINDING.  
 FT DOMAIN 238 287 PHORBOL-ESTER AND DAG BINDING.  
 FT DOMAIN 375 634 PROTEIN KINASE.  
 FT NP\_BIND 381 389 ATP (BY SIMILARITY).  
 FT BINDING 404 404 ATP (BY SIMILARITY).  
 FT ACT\_SITE 499 499 BY SIMILARITY.  
 FT MOD\_RES 89 89 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 FT MOD\_RES 139 139 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 FT MOD\_RES 324 324 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 FT VARSPLOC 1 137 MISSING (IN TPA-1B).  
 SQ SEQUENCE 704 AA; 80298 MW; 84720228 CRC32;

Query Match 78.1%; Score 50; DB 1; Length 704;  
 Best Local Similarity 70.0%; Pred. No. 7.16e-01;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 127 APDDHPVED 136  
 QY 1 APPEDNPVED 10  
 |||||

RESULT 3  
 ID LCB2.KLULA STANDARD; PRT; 562 AA.  
 AC P48241;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE SERINE PALMITOYLTRANSFERASE 2 (EC 2.3.1.50) (LONG CHAIN BASE  
 DE BIOSYNTHESIS PROTEIN 2) (SPT 2).  
 GN LCB2.  
 OS KLUYVEROMYCES LACTIS (YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
 OC SACCHAROMYCETACEAE; KLUYVEROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JA6;  
 RX MEDLINE; 97080528.  
 RA NAGIEC M.M., LESTER R.L., DICKSON R.C.;  
 RT "Sphingolipid synthesis: identification and characterization of  
 RT mammalian cDNAs encoding the Lcb2 subunit of serine  
 RT palmitoyltransferase.";  
 RL GENE 177:237-241(1996).  
 CC -!- CATALYTIC ACTIVITY: PALMITOYL-COA + L-SERINE - COA + 3-DEHYDRO-  
 CC D-SPHINGANINE + CO(2).  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF THE LONG-CHAIN BASE  
 CC COMPONENT OF SPHINGOLIPIDS.  
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.  
 CC -!- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U15646; G1001949; -  
 CC DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; 1.  
 DR PFAM; PF00222; aminotran\_2; 1.  
 KW TRANSFERASE; ACYLTRANSFERASE; TRANSMEMBRANE; PYRIDOXAL PHOSPHATE.  
 FT TRANSMEM 61 81 POTENTIAL.  
 FT TRANSMEM 355 375 POTENTIAL.  
 FT BINDING 365 365 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 562 AA; 63004 MW; 901ACF38 CRC32;



```

Query Match      71.9%  Score 46;  DB 1;  Length 562;
Best Local Similarity 77.8%  Pred. No. 5.64e+00;
Matches          7;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;

Db 13 PPDIPLD 21
   11111111
QY 2 PPEDNPV 10

RESULT 4
ID TRBP_MOUSE STANDARD; PRT; 365 AA.
AC P97473;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PRORAMINE-1 RNA BINDING PROTEIN (PRM-1 RNA BINDING PROTEIN).
GN TARBP2 OR PRBP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE-TESTIS;
RX MEDLINE; 96220490.
RA LEE K., FAJARDO M.A., BRAUN R.E.;
RT "A testis cytoplasmic RNA-binding protein that has the properties of
a translational repressor."
RL MOL. CELL. BIOL. 16:3023-3034(1996).
CC -!- FUNCTION: BINDS IN VITRO TO THE PRM1 3'UTR. SEEMS TO ACT AS A
GENERAL REPRESSOR OF TRANSLATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC -!- SIMILARITY: CONTAINS 3 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U79962; G1737216;
DR MGD; MGI:108201; PRBP.
DR PFAM; PF00035; dsrm; 2.
DR HSSP; P25159; LSTU.
KW RNA-BINDING; REPEAT; NUCLEAR PROTEIN.
FT DOMAIN 79 95
FT DOMAIN 208 224 DRBM 1.
FT DOMAIN 342 358 DRBM 2.
FT DOMAIN 342 358 DRBM 3.
SQ SEQUENCE 365 AA; 38789 MW; 0AF5653D CRC32;

Query Match      70.3%  Score 45;  DB 1;  Length 365;
Best Local Similarity 75.0%  Pred. No. 9.28e+00;
Matches          6;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;

Db 117 SPEDTPV 124
   11111111
QY 1 APPEDNPV 8

RESULT 5
ID RB_HUMAN STANDARD; PRT; 928 AA.
AC P06400; P78499;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RETINOBLASTOMA-ASSOCIATED PROTEIN (PPI10) (P105-RB) (RB).
GN RB1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE; 88014238.
RA LEE W.-H., SHEW J.-Y., HONG F.D., SERY T.W., DONOSO L.A., YOUNG L.-J.,
RA BOOKSTEIN R., LEE E.Y.-H.P.;
RT "The retinoblastoma susceptibility gene encodes a nuclear
phosphoprotein associated with DNA binding activity."
RL NATURE 329:642-645(1987).
RN [2]
RP REVISIONS.
RX MEDLINE; 87149066.
RA LEE W.-H., BOOKSTEIN R., HONG F.D., YOUNG L.-J., SHEW J.-Y.,
RA LEE E.Y.-H.P.;
RT "Human retinoblastoma susceptibility gene: cloning, identification,
and sequence."
RL SCIENCE 235:1394-1399(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88097427.
RA FRIEND S.H., HOROWITZ J.M., GERBER M.R., WANG X.-F., BOGENMANN E.,
RA LI F.P., WEINBERG R.A.;
RT "Deletions of a DNA sequence in retinoblastomas and mesenchymal
tumors: organization of the sequence and its encoded protein."
RL PROC. NATL. ACAD. SCI. U.S.A. 84:9059-9063(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90006771.
RA MCGEE T.D., YANDELL D.W., DRYJA T.P.;
RT "Structure and partial genomic sequence of the human retinoblastoma
susceptibility gene."
RL GENE 80:119-128(1989).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94063891.
RA TOGUCHIDA J., MCGEE T.L., CICCARELLI J.A., EAGLE J.R., YANDELL D.W.,
RA DRYJA T.P.;
RT "Complete genomic sequence of the human retinoblastoma susceptibility
gene."
RL GENOMICS 17:535-543(1993).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92319557.
RA HOGG A., ONADIM Z., BAIRD P.N., CONELL J.K.;
RT "Detection of heterozygous mutations in the Rb1 gene in
retinoblastoma patients using single-strand conformation polymorphism
analysis and polymerase chain reaction sequencing."
RL ONCOGENE 7:1445-1451(1992).
RN [7]
RP SEQUENCE OF 1-45 FROM N.A.
RX MEDLINE; 89239464.
RA T'ANG A., WU K.J., HASHIMOTO T., LIU W.Y., TAKAHASHI R., SHI X.H.,
RA MIHARA K., ZHANG F.H., CHEN Y.Y., DU C., QIAN J., LIN Y.G.,
RA MURPHREE A.L., QIU W.R., THOMPSON T., BENEDICT W.F., FUNG Y.K.T.;
RT "Genomic organization of the human retinoblastoma gene."
RL ONCOGENE 4:401-407(1989).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 378-562.
RX MEDLINE; 97290453.
RA KIM H.Y., CHO Y.;
RT "Structural similarity between the pocket region of retinoblastoma
tumour suppressor and the cyclin-box."
RL NAT. STRUCT. BIOL. 4:390-395(1997).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 380-785.
RX MEDLINE; 98154728.
RA LEE J.O., RUSSO A.A., PAVLETICH N.P.;
RT "Structure of the retinoblastoma tumour-suppressor pocket domain bound
to a peptide from HPV E7."
RL NATURE 391:859-865(1998).
RN [10]
RP PHOSPHORYLATION SITES.
RX MEDLINE; 92097348.
RA LEES J.A., BUCHKOVICH K.J., MARSHAK D.R., ANDERSON C.W., HARLOW E.;

```

"The retinoblastoma protein is phosphorylated on multiple sites by human cdc2";  
EMBO J. 10:4279-4290(1991).  
[11]  
RN VARIANT RB LEU-567.  
RX MEDLINE: 90081757.  
RA YANDELL D.W., CAMPBELL T.A., DAYTON S.H., PETERSEN R., WALTON D.,  
RA LITTLE J.B., MCCONKIE-ROSELL A., BUCKLEY E., DRYJA T.;  
RT "Oncogenic point mutations in the human retinoblastoma gene: their  
RT application to genetic counseling.";  
RL NEW ENGL. J. MED. 321:1689-1695(1989).  
[12]  
RN VARIANT RB TRP-661.  
RX MEDLINE: 92335261.  
RA ONADIM Z., HOGG A., BAIRD P.N., COWELL J.K.;  
RT "Oncogenic point mutations in exon 20 of the Rb1 gene in families  
RT showing incomplete penetrance and mild expression of the  
RT retinoblastoma phenotype.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 89:6177-6181(1992).  
[13]  
RN VARIANT RB ARG-457.  
RX MEDLINE: 93348271.  
RA HOGG A., BIA B., ONADIM Z., COWELL J.K.;  
RT "Molecular mechanisms of oncogenic mutations in tumors from patients  
RT with bilateral and unilateral retinoblastoma.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 90:7351-7355(1993).  
[14]  
RN VARIANTS RB ASN-480 DEL AND TRP-661.  
RX MEDLINE: 95012220.  
RA LOHMANN D.R., BRANDT B., HOEPFING W., PASSARGE E., HORSTHEMKE B.;  
RT "Distinct Rb1 gene mutations with low penetrance in hereditary  
RT retinoblastoma.";  
RL HUM. GENET. 94:349-354(1994).  
[15]  
RN VARIANTS RB GLN-72; TYR-549 AND LYS-803.  
RX MEDLINE: 96187126.  
RA LIU Z., SONG Y., BIA B., COWELL J.K.;  
RT "Germline mutations in the Rb1 gene in patients with hereditary  
RT retinoblastoma.";  
RL GENES CHROMOSOMES CANCER 14:277-284(1995).  
[16]  
RN VARIANTS RB THR-185; PRO-635; VAL-654 AND PRO-685.  
RX MEDLINE: 95315934.  
RA BLANQUET V., TURLEAU C., GROSS-MORAND M.S., SENNAUD-BEAUFORT C.,  
RA DOZ F., BESMOND C.;  
RT "Spectrum of germline mutations in the Rb1 gene: a study of 232  
RT patients with hereditary and non hereditary retinoblastoma.";  
RL HUM. MOL. GENET. 4:383-388(1995).  
[17]  
RN VARIANTS RB ASP-137 AND TRP-661.  
RX MEDLINE: 97456418.  
RA LOHMANN D.R., GERICK M., BRANDT B., OELSCHLAEGER U., LORENZ B.,  
RA PASSARGE E., HORSTHEMKE B.;  
RT "Constitutional Rb1-gene mutations in patients with isolated  
RT unilateral retinoblastoma.";  
RL AM. J. HUM. GENET. 61:282-294(1997).  
[18]  
RN VARIANTS RB LEU-567; ARG-712 AND PRO-662.  
RA YILMAZ S., HORSTHEMKE B., LOHMANN D.R.;  
RT "Twelve novel Rb1 gene mutations in patients with hereditary  
RT retinoblastoma.";  
RL HUM. MUTAT. 12:434-434(1998).  
CC -1- FUNCTION: PROBABLY ACTS AS A REGULATOR OF OTHER GENES. FORMS A  
CC COMPLEX WITH ADENOVIRUS E1A AND WITH SV40 LARGE T ANTIGEN. ACTS AS  
CC A TUMOR SUPPRESSOR. MAY BIND AND MODULATE FUNCTIONALLY CERTAIN  
CC CELLULAR PROTEINS WITH WHICH T AND E1A COMPETE FOR POCKET BINDING.  
CC POTENT INHIBITOR OF E2F-MEDIATED TRANS-ACTIVATION, INTERACTS  
CC PREFERENTIALLY WITH TRANSCRIPTION FACTOR E2F1.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- PTM: PHOSPHORYLATED FROM S TO M PHASE OF THE CELL CYCLE AND IS  
CC DEPHOSPHORYLATED IN G1. T, BUT NOT E1A, BINDS ONLY TO THE  
CC UNPHOSPHORYLATED FORM  
CC -1- DISEASE: DELETION OR ALTERATION OF THIS PROTEIN RESULTS IN THE

CHILDHOOD CANCER RETINOBLASTOMA (RB). RB IS A CONGENITAL  
MALIGNANT TUMOR THAT ARISES FROM THE NUCLEAR LAYERS OF THE  
RETINA. IT OCCURS IN ABOUT 1:20,000 LIVE BIRTHS AND REPRESENTS  
ABOUT 2% OF CHILDHOOD MALIGNANCIES. IT IS BILATERAL IN ABOUT 30%  
OF CASES. ALTHOUGH MOST RB APPEAR SPORADICALLY, ABOUT 20% ARE  
TRANSMITTED AS AN AUTOSOMAL DOMINANT TRAIT WITH INCOMPLETE  
PENETRANCE. THE DIAGNOSIS IS USUALLY MADE BEFORE THE AGE OF 2  
YEARS WHEN STRABISMUS OR A GREY TO YELLOW REFLEX FROM PUPIL ("CAT  
EYE") IS INVESTIGATED.  
CC -1- DATABASE: NAME=RB1base; NOTE=Rb1 mutation db;  
CC WWW="http://home.kamp.net/home/dr.lohmann/";  
CC -1- SIMILARITY: BELONGS TO THE RETINOBLASTOMA PROTEIN (RB) FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L41870; G793995; -  
DR EMBL; M27866; G521212; -  
DR EMBL; M27845; G521212; JOINED.  
DR EMBL; M27846; G521212; JOINED.  
DR EMBL; M27847; G521212; JOINED.  
DR EMBL; M27849; G521212; JOINED.  
DR EMBL; M27850; G521212; JOINED.  
DR EMBL; M27851; G521212; JOINED.  
DR EMBL; L35146; G521212; JOINED.  
DR EMBL; M27852; G521212; JOINED.  
DR EMBL; M27853; G521212; JOINED.  
DR EMBL; M27854; G521212; JOINED.  
DR EMBL; M27855; G521212; JOINED.  
DR EMBL; M27856; G521212; JOINED.  
DR EMBL; M27857; G521212; JOINED.  
DR EMBL; M27858; G521212; JOINED.  
DR EMBL; M27859; G521212; JOINED.  
DR EMBL; M27860; G521212; JOINED.  
DR EMBL; L35147; G521212; JOINED.  
DR EMBL; M27862; G521212; JOINED.  
DR EMBL; M27863; G521212; JOINED.  
DR EMBL; M27864; G521212; JOINED.  
....  
Note: remainder of annotations omitted.  
Query Match 70.3%; Score 45; DB 1; Length 928;  
Best Local Similarity 50.0%; Pred No. 9,28e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 27 PPPEEDPEQD 36  
:||||:|  
Qy 1 APPEDNPVED 10  
RESULT 6  
ID SYL\_ARCFU STANDARD; PRT; 932 AA.  
AC O30250;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE LEUCYL-TRNA SYNTHETASE (EC 6.1.1.4) (LEUCINE--TRNA LIGASE) (LEURS).  
GN LEUS OR AF2421.  
OS ARCHAEoglobus fulgidus.  
OC ARCHAEA; EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae;  
OC ARCHAEoglobus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE: 98049343.  
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,  
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,  
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,

RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,  
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,  
 RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,  
 RA OVERBEEK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,  
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,  
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,  
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESSE C.R.,  
 RA VENTER J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon *Archaeoglobus fulgidus*.";  
 RT NATURE 390:364-370(1997).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-LEUCINE + TRNA(LEU) = AMP +  
 CC PYROPHOSPHATE + L-LEUCYL-TRNA(LEU).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AE001108; G3650668; -;  
 DR TIGR: AF2421; -;  
 DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_I; 1;  
 DR PFAM: PF00133; TRNA-synt\_1; 1;  
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.  
 FT SIMILAR 38 48 "HIGH" REGION.  
 FT SIMILAR 630 634 "KMSKS" REGION.  
 FT BINDING 633 633 ATP (BY SIMILARITY).  
 SQ SEQUENCE 932 AA; 108628 MW; 13803D54 CRC32;  
 Query Match 70.3%; Score 45; DB 1; Length 932;  
 Best Local Similarity 66.7%; Pred. No. 9.28e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 180 PHOENPVED 188  
 I:::|||||  
 QY 2 PPEDNPVED 10  
 RESULT 7  
 ID NF1\_HUMAN STANDARD; PRT; 2839 AA.  
 AC F21359;  
 DT 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).  
 GN NF1.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE OF 1-1370 AND 1392-2839 FROM N.A.  
 RX MEDLINE: 92147138.  
 RA MARCHUK D.A., SAULINO A., TAVAKKOL R., SWAROOP M., WALLACE M.R.,  
 RA ANDERSEN L.B., MITCHELL A.L., GUTMANN D.H., BOGUSKI M., COLLINS F.S.;  
 RT "cDNA cloning of the type 1 neurofibromatosis gene: complete sequence  
 RT of the NF1 gene product.";  
 RL GENOMICS 11:931-940(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 93090270.  
 RA BERNARDS A., HAASE V.H., MURTHY A.E., MENON A., HANNIGAN G.E.,  
 RA GUSELLA J.F.;  
 RT "Complete human NF1 cDNA sequence: two alternatively spliced mRNAs  
 RT and absence of expression in a neuroblastoma line.";  
 RL DNA CELL BIOL. 11:727-734(1992).  
 RN [3]  
 RP SEQUENCE OF 335-1370 AND 1392-2839 FROM N.A.  
 RX MEDLINE: 90335969.

RA XU G., O'CONNELL P., VISKOCHIL D., CANTHON R., ROBERTSON M.,  
 RA CULVER M., DUNN D., STEVENS J., GESTELAND R., WHITE R., WEISS R.;  
 RT "The neurofibromatosis type 1 gene encodes a protein related to GAP.";  
 RL CELL 62:599-608(1990).  
 RN [4]  
 RP SEQUENCE OF 1096-1370 AND 1372-1590 FROM N.A.  
 RX MEDLINE: 91029515.  
 RA MARTIN G.A., VISKOCHIL D., BOLLAG G., MCCABE P.C., CROSTER W.J.,  
 RA HAUBRUCK H., CONROY L., CLARK R., O'CONNELL P., CANTHON R.M.,  
 RA INNIS M., MCCORMICK F.;  
 RT "The GAP-related domain of the neurofibromatosis type 1 gene product  
 RT interacts with ras p21.";  
 RL CELL 63:843-849(1990).  
 RN [5]  
 RP SEQUENCE OF 1606-2709 FROM N.A., AND VARIANT PRO-1953.  
 RX MEDLINE: 90304909.  
 RA CANTHON R.M., WEISS R., XU G., VISKOCHIL D., CULVER M., STEVENS J.,  
 RA ROBERTSON M., DUNN D., GESTELAND R., O'CONNELL P., WHITE R.;  
 RT "A major segment of the neurofibromatosis type 1 gene: cDNA sequence,  
 RT genomic structure, and point mutations.";  
 RL CELL 62:193-201(1990).  
 RN [6]  
 RP SEQUENCE OF 2230-2839 FROM N.A.  
 RX MEDLINE: 90319792.  
 RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., LETCHER R., ODEH H.M.,  
 RA SAULINO A.M., FOUNTAIN J.W., BRETON A., NICHOLSON J., MITCHELL A.L.,  
 RA BROWNSTEIN B.H., COLLINS F.S.;  
 RT "Type 1 neurofibromatosis gene: identification of a large transcript  
 RT disrupted in three NF1 patients.";  
 RL SCIENCE 249:181-186(1990).  
 RN [7]  
 RP ERRATUM.  
 RX MEDLINE: 91102559.  
 RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., COLLINS F.S.;  
 RL SCIENCE 250:1749-1749(1990).  
 RN [8]  
 RP SEQUENCE OF 1168-1566 FROM N.A.  
 RX MEDLINE: 92019823.  
 RA NISHI T., LEE P.S., OKA K., LEVIN V.A., TANASE S., MORINO Y.,  
 RA SAYA H.;  
 RT "Differential expression of two types of the neurofibromatosis type 1  
 RT (NF1) gene transcripts related to neuronal differentiation.";  
 RL ONCOGENE 6:1555-1559(1991).  
 RN [9]  
 RP SEQUENCE OF 1371-1391 FROM N.A.  
 RX MEDLINE: 93109335.  
 RA ANDERSEN L.B., BALLESTER R., MARCHUK D.A., CHANG E., GUTMANN D.H.,  
 RA SAULINO A.M., CAMONIS J., WIGLER M., COLLINS F.S.;  
 RT "A conserved alternative splice in the von Recklinghausen  
 RT neurofibromatosis (NF1) gene produces two neurofibromin isoforms,  
 RT both of which have GTPase-activating protein activity.";  
 RL MOL. CELL. BIOL. 13:487-495(1993).  
 RN [10]  
 RP FUNCTION.  
 RX MEDLINE: 91029516.  
 RA BALLESTER R., MARCHUK D., BOGUSKI M.S., SAULINO A., LETCHER R.,  
 RA WIGLER M., COLLINS F.S.;  
 RT "The NF1 locus encodes a protein functionally related to mammalian  
 RT GAP and yeast IRA proteins.";  
 RL CELL 63:851-859(1990).  
 RN [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE: 95072625.  
 RA UPADHYAYA M., SHAW D.J., HARPER P.S.;  
 RT "Molecular basis of neurofibromatosis type 1 (NF1): mutation analysis  
 RT and polymorphisms in the NF1 gene.";  
 RL HUM. MUTAT. 4:83-101(1994).  
 RN [12]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE: 96422425.  
 RA HONG SHEN M., HARPER P.S., UPADHYAYA M.;  
 RT "Molecular genetics of neurofibromatosis type 1 (NF1).";  
 RL J. MED. GENET. 33:2-17(1996).



CC -!- TISSUE SPECIFICITY: TYPE I IS EXPRESSED PREDOMINANTLY IN BRAIN,  
CC SPINAL CORD AND TESTIS. TYPE II IS EXPRESSED PREDOMINANTLY IN  
CC ADRENAL GLAND, KIDNEY, OVARY AND LUNG. TYPE III IS EXPRESSED  
CC PREDOMINANTLY IN ADRENAL GLAND AND TYPE IV IS EXPRESSED  
CC MAINLY IN THE TESTIS  
CC -!- ALTERNATIVE PRODUCTS: FOUR FORMS OF THE PROTEIN (TYPES I, II,  
CC III AND IV) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME  
CC GENE. THE SEQUENCE SHOW HERE IS THAT OF TYPE II.  
CC -!- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L10369; G309451; -  
CC EMBL: L10367; G309451; JOINED.  
CC EMBL: L10368; G309451; JOINED.  
CC EMBL: L10370; G309453; -  
CC EMBL: X54924; G930191; -  
CC EMBL: D30730; G577638; -  
CC EMBL: D30731; G577640; -  
CC MGI: 97306; NE1.  
CC PROSITE: PS00509; RAS\_GTPASE\_ACTIV\_1; 1.  
CC PROSITE: PS50018; RAS\_GTPASE\_ACTIV\_2; 1.  
CC PFAM: PF00816; RASGAP; 1.  
CC GTPASE ACTIVATION; ALTERNATIVE SPLICING.  
CC KW DOMAIN 1237 1453  
CC FT VARSPLIC 1373 1393 MISSING (IN TYPE I AND TYPE IV).  
CC FT VARSPLIC 1394 1406 VVSOREPONSIGA -> VKSSCFCLNNRWLASASLRT  
CC FT VARSPLIC 1407 2841 ASVP (IN TYPE III AND TYPE IV).  
CC FT VARSPLIC 1407 2841 MISSING (IN TYPE III AND TYPE IV).  
CC SQ SEQUENCE 2841 AA; 319591 MW; A7AA76F4 CRC32;  
  
Query Match 70.3%; Score 45; DB 1; Length 2841;  
Best Local Similarity 60.08; Pred. No. 9.28e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Db 1549 GPPEHRPVAD 1558  
: : : : :  
QY 1 APPEDNPVED 10  
  
RESULT 9  
ID PAHO\_STRCA STANDARD; PRT; 36 AA.  
AC P11967;  
DT 01-OCT-1989 (REL. 12, CREATED)  
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE PANCREATIC HORMONE (PANCREATIC POLYPEPTIDE) (PP).  
OS STRUTHIO CAMELUS (OSTRICH).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC PALAEOGNATHAE; STRUTHIONIFORMES; STRUTHIONIDAE; STRUTHIO.  
RN [1]  
RP SEQUENCE.  
RX MEDLINE; 87307111.  
RA LITTHAUER D., OELOFSEN W.;  
RT "Purification and primary structure of ostrich pancreatic  
RT polypeptide."  
RL INT. J. PEPT. PROTEIN RES. 29:739-745(1987).  
CC -!- FUNCTION: PANCREATIC HORMONE IS SYNTHESIZED IN PANCREATIC ISLETS  
CC OF LANGERHANS AND ACTS AS A REGULATOR OF PANCREATIC AND  
CC GASTROINTESTINAL FUNCTIONS.  
CC -!- SIMILARITY: BELONGS TO THE PP FAMILY (PANCREATIC HORMONE,  
CC NEUROPEPTIDE Y, PEPTIDE YY).  
CC PIR: A28578; A28578.  
CC PROSITE: PS00265; PANCREATIC\_HORMONE; 1.  
CC DR PFAM: PF00159; hormone3; 1.  
CC DR HSP: P01306; 1PPT.  
CC HORMONE; AMIDATION; PANCREAS.  
KW

FT MOD\_RES 36 36 AMIDATION.  
SQ SEQUENCE 36 AA; 4209 MW; 45F8AEDB CRC32;  
  
Query Match 68.8%; Score 44; DB 1; Length 36;  
Best Local Similarity 66.7%; Pred. No. 1.51e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
Db 8 PGDDAPVED 16  
: : : : :  
QY 2 PPEDNPVED 10  
  
RESULT 10  
ID PAHO\_LARAR STANDARD; PRT; 36 AA.  
AC P41337;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE PANCREATIC HORMONE (PANCREATIC POLYPEPTIDE) (PP).  
OS LARUS ARGENTATUS (HERRING GULL).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; CHARADRIIFORMES; LARIDAE; LARUS.  
RN [1]  
RP SEQUENCE.  
RC TISSUE=PANCREAS;  
RX MEDLINE; 94229519.  
RA BARTON C.L., SHAW C., HALTON D.W., THIM L.;  
RT "Isolation and structural characterisation of herring gull (Larus  
RT argentatus) pancreatic polypeptide."  
RL GEN. COMP. ENDOCRINOL. 93:255-259(1994).  
CC -!- FUNCTION: PANCREATIC HORMONE IS SYNTHESIZED IN PANCREATIC ISLETS  
CC OF LANGERHANS AND ACTS AS A REGULATOR OF PANCREATIC AND  
CC GASTROINTESTINAL FUNCTIONS.  
CC -!- SIMILARITY: BELONGS TO THE PP FAMILY (PANCREATIC HORMONE,  
CC NEUROPEPTIDE Y, PEPTIDE YY).  
CC DR PROSITE: PS00265; PANCREATIC\_HORMONE; 1.  
CC DR PFAM: PF00159; hormone3; 1.  
CC DR HSP: P01306; 1PPT.  
CC HORMONE; AMIDATION; PANCREAS.  
KW MOD\_RES 36 36 AMIDATION.  
FT MOD\_RES 36 36 AMIDATION.  
SQ SEQUENCE 36 AA; 4237 MW; 2AF9A777 CRC32;  
  
Query Match 68.8%; Score 44; DB 1; Length 36;  
Best Local Similarity 66.7%; Pred. No. 1.51e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
Db 8 PGDDAPVED 16  
: : : : :  
QY 2 PPEDNPVED 10  
  
RESULT 11  
ID PAHO\_CHICK STANDARD; PRT; 80 AA.  
AC P01306;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE PANCREATIC HORMONE PRECURSOR (PANCREATIC POLYPEPTIDE) (PP).  
OS GALLUS GALLUS (CHICKEN), AND MELEAGRIS GALLOPAVO (COMMON TURKEY).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=CHICKEN; TISSUE=LIVER;  
RX MEDLINE; 93366173.  
RA NATA K., SUGIMOTO T., KOHRI K., HIDAKA H., HATTORI E., YAMAMOTO H.,  
RA YONEKURA H., OKAMOTO H.;  
RT "Structure determination and evolution of the chicken cDNA and gene  
RT encoding prepropancreatic polypeptide."  
RL GENE 130:183-189(1993).  
RN [2]  
RP SEQUENCE OF 26-61.  
RC SPECIES=CHICKEN;

```

RX MEDLINE; 76069270.
RA KIMMEL J.R., HAYDEN L.J., POLLOCK H.G.;
RT "Isolation and characterization of a new pancreatic polypeptide
RN hormone."
RL J. BIOL. CHEM. 250:9369-9376(1975).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).
RC SPECIES=M.GALLOPAGO;
RX MEDLINE; 84179397.
RA GLOVER I., MANEEL I., PITTS J., WOODS S., MOSS D., TICKLE I.,
RA BLONDELL T.L.;
RT "Conformational flexibility in a small globular hormone: X-ray
RT analysis of avian pancreatic polypeptide at 0.98-A resolution."
RL BIOPOLYMERS 22:293-304(1983).
CC -!- FUNCTION: PANCREATIC HORMONE IS SYNTHESIZED IN PANCREATIC ISLETS
CC OF LANGERHANS AND ACTS AS A REGULATOR OF PANCREATIC AND
CC GASTROINTESTINAL FUNCTIONS.
CC -!- SIMILARITY: BELONGS TO THE PP FAMILY (PANCREATIC HORMONE,
CC NEUROPEPTIDE Y, PEPTIDE YY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13761; G391634; -
DR EMBL; D13760; G391646; -
DR PIR; A01575; PCCH.
DR PDB; 1PPT; 15-OCT-91.
DR PROSITE; PS00265; PANCREATIC_HORMONE; 1.
DR PFAM; PF00159; Hormones; 1.
KW HORMONE; CLEAVAGE ON PAIR OF BASIC RESIDUES; PANCREAS; SIGNAL;
KW AMIDATION; 3D-STRUCTURE.
FT SIGNAL 1 25
FT CHAIN 26 61
FT MOD_RES 61 61
FT CONFLICT 47 48
FT TURN 35 36
FT HELIX 39 56
FT TURN 57 58
SQ SEQUENCE 80 AA; 8773 MW; DE642B29 CRC32;

Query Match 68.8%; Score 44; DB 1; Length 80;
Best Local Similarity 66.7%; Pred. No. 1.51e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 33 PGDDAPVED 41
Qy 2 PPEDNPVED 10

RESULT 12
ID CLP1_HELAM STANDARD; PRT; 109 AA.
AC Q02443;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE LARVAL CUTICLE PROTEIN 1 PRECURSOR.
GN LCPI
OS HELICOVERPA ARMIGERA (COTTON BOLLWORM).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; LEPIDOPTERA; NOCTUIDAE; NOCTUIDAE; HELICOVERPA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTEGUMENT;
RA HENRY S.M., HOBBS A.A.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: COMPONENT OF THE CUTICLE OF THE LARVA OF HELICOVERPA
CC ARMIGERA.
CC -!- SIMILARITY: CONTAINS A CUTICLE CONSENSUS DOMAIN.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AF004945; G2209362; -
DR PROSITE; PS00233; CUTICLE; 1.
DR PFAM; PF00379; Insect_cuticle; 1.
KW STRUCTURAL PROTEIN; CUTICLE; SIGNAL.
FT SIGNAL 1 14
FT CHAIN 15 109
FT DOMAIN 68 71
FT POLY-VAL.
SQ SEQUENCE 109 AA; 11856 MW; 80888G93 CRC32;

Query Match 68.8%; Score 44; DB 1; Length 109;
Best Local Similarity 55.6%; Pred. No. 1.51e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 15 APPAEVPQ 23
Qy 1 APPEDNPVE 9

RESULT 13
ID YLC3_YEREN STANDARD; PRT; 122 AA.
AC P21208;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 13.7 KD PROTEIN IN LCRC 3' REGION (ORF3).
OS YERSINIA ENTEROCOLITICA.
OG PLASMID PV.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC YERSINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE O:3;
RX MEDLINE; 90264308.
RA VIITANEN A.-M., TOIVANEN P., SKURNIK M.;
RT "The lcr gene is part of an operon in the lcr region of Yersinia
RT enterocolitica O:3."
RL J. BACTERIOL. 172:3152-3162(1990).
CC -!- INDUCTION: TEMPERATURE SEEMS TO PLAY THE MAJOR ROLE IN REGULATION
CC OF TRANSCRIPTION OF THE LCRC-CONTAINING OPERON OF PV. WHEREAS
CC CA(2+) CONCENTRATION HAS ONLY A MODERATE EFFECT AT 37 DEGREES
CC CELSIUS, AND NO EFFECT AT ROOM TEMPERATURE.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M32097; -; NOT_ANNOTATED_CDS.
DR PIR; D35392; D35392.
KW HYPOTHETICAL PROTEIN; PLASMID.
SQ SEQUENCE 122 AA; 13741 MW; 0E8B386D CRC32;

Query Match 68.8%; Score 44; DB 1; Length 122;
Best Local Similarity 62.5%; Pred. No. 1.51e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 34 PPDGHVPE 41
Qy 2 PPEDNPVE 9
```

```
RESULT 14
ID NEUM_CARAU STANDARD; PRT; 213 AA.
AC P17691;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE NEUROMODULIN (AXONAL MEMBRANE PROTEIN GAP-43) (PP46) (B-50) (PROTEIN
F1) (CALMODULIN-BINDING PROTEIN P-57).
OS CARASSIUS AURATUS (GOLDFISH).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC CYPRINIDAE; CYPRININAE; CARASSIUS.
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE; 90380372.
RA LABATE M.E., SKENE J.H.P.;
RT "Selective conservation of GAP-43 structure in vertebrate evolution.";
RL NEURON 3:299-310(1989).
CC -!- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH NERVE GROWTH. IT IS A
CC MAJOR COMPONENT OF THE MOTILE "GROWTH CONES" THAT FORM THE TIPS
CC OF ELONGATING AXONS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF GROWTH CONE AND
CC SYNAPTIC PLASMA MEMBRANES.
CC -!- PTM: PHOSPHORYLATION OF THIS PROTEIN BY A PROTEIN KINASE C IS
CC SPECIFICALLY CORRELATED WITH CERTAIN FORMS OF SYNAPTIC PLASTICITY.
CC -!- BINDS CALMODULIN WITH A GREATER AFFINITY IN THE ABSENCE OF CA++
CC THAN IN ITS PRESENCE.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M26250; G212954;
CC PIR; JQ0075; JQ0075.
CC PROSITE; PS00412; NEUROMODULIN_1; 1.
CC PROSITE; PS00413; NEUROMODULIN_2; 1.
CC PFAM; PF00612; IQ; 1.
CC KW NEURONE; PHOSPHORYLATION; MEMBRANE; GROWTH REGULATION;
CC CALMODULIN-BINDING; LIPOPROTEIN.
CC FT DOMAIN 1 4
CC FT LIPID 3 3 PALMITATE (PROBABLE).
CC FT LIPID 4 4 PALMITATE (PROBABLE).
CC FT MOD_RES 42 42 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
CC FT DOMAIN 44 52 CALMODULIN-BINDING (PROBABLE).
CC FT DOMAIN 155 170 POLY-GLU.
CC SEQUENCE 213 AA; 23569 MW; BFAC407B CRC32;

Query Match 68.8%; Score 44; DB 1; Length 213;
Best Local Similarity 55.6%; Pred. No. 1.51e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 79 SPSEKPRVE 87
:|:|:|
QY 1 APPENDPVE 9

RESULT 15
ID TRT3_CHICK STANDARD; PRT; 262 AA.
AC P12620; P12621; P12619; P12618;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE TROPONIN T, FAST SKELETAL MUSCLE ISOFORMS.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE; 89066672.
```

```
RA SMILLIE L.B., GOLOSINSKA K., REINACH F.C.;
RT "Sequences of complete cDNAs encoding four variants of chicken
skeletal muscle troponin T.";
RL J. BIOL. CHEM. 263:18816-18820(1988).
CC -!- FUNCTION: TROPONIN T IS THE TROPOMYOSIN-BINDING SUBUNIT OF
CC TROPONIN, THE THIN FILAMENT REGULATORY COMPLEX WHICH CONFERS
CC CALCIUM-SENSITIVITY TO STRIATED MUSCLE ACTOMYOSIN ATPASE ACTIVITY.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST FOUR DIFFERENT ISOFORMS (TNT-1 TO
CC TNT-4) ARE GENERATED BY ALTERNATIVE SPLICING. THE FORM SHOWN
CC HERE IS TNT-3.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M22156; G212790;
CC EMBL; M22158; G212792;
CC EMBL; M22155; G212788;
CC EMBL; M22154; G212786;
CC PIR; A31957; A31957.
CC PIR; B31957; B31957.
CC PIR; C31957; C31957.
CC PIR; D31957; D31957.
CC PFAM; PF00992; Troponin; 1.
CC KW MUSCLE PROTEIN; MULTIGENE FAMILY; ALTERNATIVE SPLICING;
CC PHOSPHORYLATION.
CC FT INIT_MET 0 1
CC FT MOD_RES 1 1
CC FT VARSPLIC 11 34
CC FT VARSPLIC 11 16
CC FT VARSPLIC 38 43
CC FT VARSPLIC 235 247
CC FT VARNRIDAQKH -> LTLRCRLQELSKF (IN TNT-
CC 2).
CC SEQUENCE 262 AA; 31010 MW; DBC80522 CRC32;

Query Match 68.8%; Score 44; DB 1; Length 262;
Best Local Similarity 50.0%; Pred. No. 1.51e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 33 PPPEEAPPEE 42
:|:|:|
QY 1 APPENDPVED 10

Search completed: Thu Oct 21 15:32:33 1999
Job time : 7 secs.
```

\*\*\*\*\*  
MISRE  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:32:51 1999; MasPar time 4.34 Seconds  
125.894 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-040-485-6  
Description: (1-10) from US09040485.pep  
Perfect Score: 64  
Sequence: 1 APPEDNPVED 10

Scoring table: PAM 150  
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb19  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 21.084; Variance 25.523; scale 0.826

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	49	76.6	225	11	Q61167 APC-BINDING PROTEIN EB	2.78e+00
2	48	75.0	140	11	Q70201 TIAP.	4.58e+00
3	48	75.0	1175	4	Q13632 NPAT.	4.58e+00
4	48	75.0	1427	4	Q16380 NPAT (E14 AND A-T PROT	4.58e+00
5	48	75.0	1427	4	Q14207 E14 PROTEIN.	4.58e+00
6	47	73.4	129	2	Q50383 HYPOTHETICAL 13.0 KD P	7.51e+00
7	46	71.9	142	4	Q15392 APOPTOSIS INHIBITOR SU	1.22e+01
8	46	71.9	146	10	Q65828 GLUTATHIONE S-TRANSFER	1.22e+01
9	45	71.9	215	10	Q82451 GLUTATHIONE S-TRANSFER	1.22e+01
10	46	71.9	1035	10	Q81129 PHOSPHATIDYLINOSITOL	1.22e+01
11	46	71.9	1638	2	Q87001 CHPA.	1.22e+01
12	46	71.9	2042	5	Q25766 3D7VAR1 (FRAGMENT).	1.22e+01
13	45	70.3	53	4	Q92728 RETINOBLASTOMA SUSCEPT	1.98e+01
14	45	70.3	86	2	Q74580 HYPOTHETICAL 9.9 KD PR	1.98e+01
15	45	70.3	97	5	Q26168 DNA FRAGMENT WITH HIGH	1.98e+01
16	45	70.3	258	2	Q33339 DEHYDROGENASE.	1.98e+01
17	45	70.3	287	5	Q16657 CO3H5.3 PROTEIN.	1.98e+01
18	45	70.3	317	4	Q15319 RNA POLYMERASE III SUB	1.98e+01
19	45	70.3	339	2	Q56796 JUGLANDIS COPPER-RESIS	1.98e+01
20	45	70.3	840	4	Q14284 NEUROFIBROMATOSIS 1 (F	1.98e+01

21	45	70.3	1294	4	O14529 KIAA0293 (FRAGMENT).	1.98e+01
22	45	70.3	1426	11	P70298 CUT-LIKE 2 (CUX-2).	1.98e+01
23	45	70.3	2172	5	Q09515 HYPOTHETICAL 236.2 KD	1.98e+01
24	45	70.3	2820	11	P97526 NEUROFIBROMIN.	1.98e+01
25	44	68.8	122	2	Q68696 HYPOTHETICAL 13.8 KD P	3.17e+01
26	44	68.8	131	1	Q49612 FORMYLMETHANOFURAN DEH	3.17e+01
27	44	68.8	259	1	Q29756 CONSERVED HYPOTHETICAL	3.17e+01
28	44	68.8	273	13	Q91257 LMPX OF LAMPREY (EC 3.	3.17e+01
29	44	68.8	304	2	Q60225 BETA-LACTAMASE.	3.17e+01
30	44	68.8	334	10	Q24272 ACTIN 7 (FRAGMENT).	3.17e+01
31	44	68.8	348	13	Q57378 RECEPTOR-ASSOCIATED PR	3.17e+01
32	44	68.8	365	4	Q60509 NSP-LIKE 1 (FRAGMENT).	3.17e+01
33	44	68.8	461	3	Q74679 A/G-SPECIFIC ADENINE D	3.17e+01
34	44	68.8	473	5	Q26236 BETA-TUBULIN.	3.17e+01
35	44	68.8	545	4	Q75298 RTM2-A.	3.17e+01
36	44	68.8	598	2	Q50866 KEFC.	3.17e+01
37	44	68.8	739	2	Q59557 T-CATALASE (EC 1.11.1.1.	3.17e+01
38	44	68.8	740	2	Q05763 CATALASE-PEROXIDASE.	3.17e+01
39	44	68.8	1262	5	Q20684 F52H3.7 PROTEIN.	3.17e+01
40	44	68.8	1506	13	P79927 INTEGRATARY MUCIN B.	3.17e+01
41	44	68.8	2764	5	Q01399 NEUROFIBROMIN.	3.17e+01
42	44	68.8	2802	5	Q01398 NEUROFIBROMIN.	3.17e+01
43	44	68.8	2802	5	Q01397 NEUROFIBROMIN.	3.17e+01
44	44	68.8	3413	2	Q54593 POLYKETIDE SYNTHASE.	3.17e+01
45	43	67.2	507	14	O90316 GAG.	5.05e+01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	225 AA.
ID Q61167			
AC Q61167			
DT 01-NOV-1996	(TREMBLREL. 01, CREATED)		
DT 01-NOV-1996	(TREMBLREL. 01, LAST SEQUENCE UPDATE)		
DT 01-NOV-1996	(TREMBLREL. 01, LAST ANNOTATION UPDATE)		
DE APC-BINDING PROTEIN EB2 (FRAGMENT).			
GN EB2.			
OS MUS MUSCULUS (MOUSE).			
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
[1]			
RN RP	SEQUENCE FROM N.A.		
RA HOFFMAN N.G., KAY B.K.;			
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR EMBL; U51204; G1256436; -.			
FT NON_TER 1 1			
SQ SEQUENCE 225 AA; 25678 MW; 8125B8B4 CRC32;			

Query Match 76.6%; Score 49; DB 11; Length 225;  
Best Local Similarity 60.0%; Pred. No. 2.78e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 207 APPEDDEIEE 216	PRELIMINARY;	PRT;	140 AA.
QY 1 APPEDNPVED 10			
RESULT 2			
ID Q70201			
AC Q70201			
DT 01-AUG-1998	(TREMBLREL. 07, CREATED)		
DT 01-AUG-1998	(TREMBLREL. 07, LAST SEQUENCE UPDATE)		
DT 01-AUG-1998	(TREMBLREL. 07, LAST ANNOTATION UPDATE)		
DE TIAP.			
GN TIAP.			
OS MUS MUSCULUS (MOUSE).			
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
[1]			
RN RP	SEQUENCE FROM N.A.		
RC TISSUE-EMBRYO;			
RA KOBAYASHI K., OTAKI M., OGASAWARA T., TOKUHISA T.;			
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			



DR EMBL; AB013819; D1029206; -  
SQ SEQUENCE 140 AA; 16297 MW; 2BD48871 CRC32;

Query Match 75.0%; Score 48; DB 11; Length 140;  
Best Local Similarity 62.5%; Pred. No. 4.58e+00;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 69 PDDNPIEE 76  
1:||||:|  
Qy 3 PEDNPVED 10

RESULT 3  
ID Q13632 PRELIMINARY; PRT: 1175 AA.

AC Q13632;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DI 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE NPAT.

GN NPAT.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

OC CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RA CHEN X., YANG L., UDAR N., LIANG T., XU S., UHRHAMMER N., BAY J.O.,

RA WANG Z., DANDAKAR U., CHIPLUNKAR S., KLISAK I., TELATAR M., YANG H.,

RA CONCANNON P., GATTI R.A.;

RL MAMM. GENOME 0:0-0(0).

DR EMBL; U58852; G1381667; -

SQ SEQUENCE 1175 AA; 126731 MW; CFC49BD6 CRC32;

Query Match 75.0%; Score 48; DB 4; Length 1175;  
Best Local Similarity 60.0%; Pred. No. 4.58e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 688 TPPECTPVEN 697  
:|||||:  
Qy 1 APPEDNPVED 10

RESULT 4

ID Q16580 PRELIMINARY; PRT: 1427 AA.

AC Q16580;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE NPAT (E14 AND A-T PROTEINS).

GN E14 OR NPAT OR E14/NPAT.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

OC CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 97349105.

RA IMAI T., SUGAWARA T., NISHIYAMA A., SHIMADA R., OHKI R., SEKI N.,

RA SAGARA M., ITO H., YAMAUCHI M., HORI T.;

RT "The structure and organization of the human NPAT gene.";

RL GENOMICS 42:388-392(1997).

RN [2]

RP SEQUENCE OF 1-12 FROM N.A.

RA BYRD P.J., MCCONVILLE C.M., COOPER P.R., PARKHILL J., MCGUIRE G.,

RA STANKOVIC T., THICK J., TAYLOR A.M.R.;

RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]

RP SEQUENCE OF 1-12 FROM N.A.

RX MEDLINE; 96338579.

RA IMAI T., YAMAUCHI M., SEKI N., SUGAWARA T., SAITO T., MATSUDA Y.,

RA ITOH H., NAGASE T., NOMURA N., HORI T.;

RT "Identification and characterization of a new gene physically linked

to the ATM gene.";

RL GENOME RES. 6:439-447(1996).

RN [4]

RP SEQUENCE OF 1-12 FROM N.A.

RX MEDLINE; 97343327

RA PLATZER M., ROTMAN G., BAUER D., UZIEL T., SAVITSKY K., BAR-SHIRA A.,

RA GILAD S., SHILOH Y., ROSENTHAL A.;

RT "Ataxia-telangiectasia locus: sequence analysis of 184 kb of human

genomic DNA containing the entire ATM gene.";

RL GENOME RES. 7:592-605(1997).

RN [5]

RP SEQUENCE OF 1-12 FROM N.A.

RA PLATZER M., ROTMAN G., BAUER D., SAVITSKY K., SHILOH Y.,

RA ROSENTHAL A.;

RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; D89854; D1022213; -

DR EMBL; D83244; D1022213; JOINED.

DR EMBL; D89841; D1022213; JOINED.

DR EMBL; D89842; D1022213; JOINED.

DR EMBL; D89843; D1022213; JOINED.

DR EMBL; D89844; D1022213; JOINED.

DR EMBL; D89845; D1022213; JOINED.

DR EMBL; D89846; D1022213; JOINED.

DR EMBL; D89847; D1022213; JOINED.

DR EMBL; D89848; D1022213; JOINED.

DR EMBL; D89849; D1022213; JOINED.

DR EMBL; D89850; D1022213; JOINED.

DR EMBL; D89851; D1022213; JOINED.

DR EMBL; D89852; D1022213; JOINED.

DR EMBL; D89853; D1022213; JOINED.

DR EMBL; X91196; E198282; -

DR EMBL; D83244; G1304116; -

DR EMBL; U82828; G2304972; -

SQ SEQUENCE 1427 AA; 154303 MW; A7244B3F CRC32;

Query Match 75.0%; Score 48; DB 4; Length 1427;  
Best Local Similarity 60.0%; Pred. No. 4.58e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 688 TPPECTPVEN 697  
:|||||:  
Qy 1 APPEDNPVED 10

RESULT 5

ID Q14207 PRELIMINARY; PRT: 1427 AA.

AC Q14207; Q14967;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE E14 PROTEIN.

GN E14.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

OC CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RA COOPER P.R., BYRD P.J., TAYLOR A.M.R.;

RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RA BYRD P.J., MCCONVILLE C.M., COOPER P.R., PARKHILL J., MCGUIRE G.,

RA STANKOVIC T., THICK J., TAYLOR A.M.R.;

RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE-PLACENTA, AND TESTIS;

RX MEDLINE; 96338579.

RA IMAI T., YAMAUCHI M., SEKI N., SUGAWARA T., SAITO T., MATSUDA Y.,

RA ITOH H., NAGASE T., NOMURA N., HORI T.;

RT "Identification and characterization of a new gene physically linked

to the ATM gene.";

RL GENOME RES. 6:439-447(1996).

DR EMBL; X97186; E238806; -

DR EMBL; D83243; G1304114; -

FT CONFLICT 14 14 Y -> N (IN REF. 2).

FT CONFLICT 295 295 L -> I (IN REF. 3).  
 FT CONFLICT 471 471 N -> Y (IN REF. 3).  
 FT CONFLICT 967 967 E -> Q (IN REF. 3).  
 FT CONFLICT 973 973 V -> L (IN REF. 3).  
 FT CONFLICT 987 987 A -> V (IN REF. 3).  
 SQ SEQUENCE 1427 AA; 154273 MW; 2DDEBF4 CRC32;

Query Match 75.0%; Score 48; DB 4; Length 1427;  
 Best Local Similarity 60.0%; Pred. No. 4.58e+00;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 688 TPPEGTPTVEN 697  
 :||| |||  
 QY 1 APPEDNPVED 10

RESULT 6  
 ID O50383 PRELIMINARY; PRT; 129 AA.  
 AC O50383;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 13.0 KD PROTEIN.  
 GN MT0004.11.  
 OS MYCOBACTERIUM TUBERCULOSIS.  
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
 OC ACTINOMYCETALES; CORINEBACTERIINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RA OLIVER K., SKELTON J., BADCOCK K., CHURCHER C.M., HARRIS D.;  
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.;  
 RL "An integrated map of the genome of the tubercle bacillus,  
 RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium  
 RT leprae.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137 (1996).  
 DR EMBL; AL009198; E1202271; -.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 129 AA; 12988 MW; 17B14A83 CRC32;

Query Match 73.4%; Score 47; DB 2; Length 129;  
 Best Local Similarity 60.0%; Pred. No. 7.51e+00;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 28 APAQNPVDD 37  
 :||| |||  
 QY 1 APPEDNPVED 10

RESULT 7  
 ID O15392 PRELIMINARY; PRT; 142 AA.  
 AC O15392;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE APOPTOSIS INHIBITOR SURVIVIN.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE; 94148797.  
 RA ALTIERI D.C.;  
 RT "Molecular cloning of effector cell protease receptor-1, a novel cell  
 RT surface receptor for the protease factor xa.";  
 RL J. BIOL. CHEM. 269:3139-3142(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95034823.  
 RA ALTIERI D.C.;  
 RT "Splicing of effector cell protease receptor-1 mRNA is modulated by  
 RT an unusual retained intron.";  
 RL BIOCHEMISTRY 33:13848-13855(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97398388.  
 RA AMBROSINI G., ADIDA C., ALTIERI D.C.;  
 RT "A novel anti-apoptosis gene, survivin, expressed in cancer and  
 RT lymphoma.";  
 RL NAT. MED. 3:917-921(1997).  
 DR EMBL; U75285; G2315863; -.  
 DR PFAM; PF00653; BIR; 1.  
 SQ SEQUENCE 142 AA; 16389 MW; 702152A2 CRC32;

Query Match 71.9%; Score 46; DB 4; Length 142;  
 Best Local Similarity 50.0%; Pred. No. 1.22e+01;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 69 PDDPIEE 76  
 :||: |||  
 QY 3 PEDNPVED 10

RESULT 8  
 ID O65858 PRELIMINARY; PRT; 146 AA.  
 AC O65858;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE GLUTATHIONE S-TRANSFERASE (FRAGMENT).  
 GN RGST II.  
 OS ORYZA SATIVA (RICE).  
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;  
 OC POACEAE; ORYZA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. LEMONT;  
 RA WU J., CRAMER C.L., HATZIOS K.K.;  
 RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AJ002381; E1289811; -.  
 KW TRANSFERASE.  
 FT NON\_TER 1  
 FT NON\_TER 146  
 FT NON\_TER 146  
 SQ SEQUENCE 146 AA; 18443 MW; AC010F6D CRC32;

Query Match 71.9%; Score 46; DB 10; Length 146;  
 Best Local Similarity 50.0%; Pred. No. 1.22e+01;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 77 APDEKVVEE 86  
 :||: |||  
 QY 1 APPEDNPVED 10

RESULT 9  
 ID O82451 PRELIMINARY; PRT; 215 AA.  
 AC O82451;  
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUTATHIONE S-TRANSFERASE II.  
 OS ORYZA SATIVA (RICE).  
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;

OC EUPHYLLIOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=ROOT;  
RA WU J., CRAMER C., HATZIOS K.K.;  
RT "Isolation of a full-length cDNA encoding the second glutathione  
S-transferase from rice (Accession No. AF062403) (PGR98-136).";  
RL PLANT PHYSIOL. 118:329-329(1998).  
DR EMBL; AF062403; G3746581; -;  
KW TRANSFERASE.  
SQ SEQUENCE 215 AA; 23977 MW; 4F6B3595 CRC32;  
Query Match 71.9%; Score 46; DB 10; Length 215;  
Best Local Similarity 50.0%; Pred. No. 1.22e+01;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 126 APPDEKVVEE 135  
|:::|:  
QY 1 APPDNPVED 10  
RESULT 10  
ID O81129 PRELIMINARY; PRT; 1035 AA.  
AC O81129;  
DT 01-NOV-1998 (TREMREL. 08, CREATED)  
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE PHOSPHATIDYLINOSITOL 4-KINASE (FRAGMENT).  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;  
OC EUPHYLLIOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, COLUMBIA;  
RX MEDLINE; 98380505;  
RA STEVENSON J.M., PERERA I.Y., BOSS W.F.;  
RT "A Phosphatidylinositol 4-Kinase Pleckstrin Homology Domain That  
Binds Phosphatidylinositol 4-Monophosphate.";  
RL J. BIOL. CHEM. 273:22761-22767(1998).  
DR EMBL; AF035936; G3452263; -;  
FT NON\_TER 1  
SQ SEQUENCE 1035 AA; 115699 MW; 7F294997 CRC32;  
Query Match 71.9%; Score 46; DB 10; Length 1035;  
Best Local Similarity 55.6%; Pred. No. 1.22e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 273 PPSPDPVQ 281  
|:::|:  
QY 2 PPDNPVED 10  
RESULT 11  
ID O87001 PRELIMINARY; PRT; 1638 AA.  
AC O87001;  
DT 01-NOV-1998 (TREMREL. 08, CREATED)  
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE CHPA.  
GN CHPA.  
OS PSEUDOMONAS AERUGINOSA.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;  
OC PSEUDOMONAS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PAOI;  
RA WHITCHURCH C.B., YOUNG M.D., HOBBS M., MATTICK J.S.;  
RT "Pseudomonas aeruginosa chemotactic transduction genes pill, chpA  
chpB and downstream genes chpC, chpD and chpE";  
PL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
EMBL; U79580; G3241969; -;

SQ SEQUENCE 1638 AA; 178594 MW; B10FD714 CRC32;  
Query Match 71.9%; Score 46; DB 2; Length 1638;  
Best Local Similarity 30.0%; Pred. No. 1.22e+01;  
Matches 3; Conservative 7; Mismatches 0; Indels 0; Gaps 0;  
Db 22 PPADPEPVDE 31  
|:|:::|:  
QY 1 APPDNPVED 10  
RESULT 12  
ID Q25766 PRELIMINARY; PRT; 2042 AA.  
AC Q25766;  
DT 01-NOV-1996 (TREMREL. 01, CREATED)  
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE 3D7VARI (FRAGMENT).  
GN 3D7VARI.  
OS PLASMIDIUM FALCIPARUM.  
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMIDIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96324414.  
RA RUBIO J.P., THOMPSON J.K., COWMAN A.F.;  
RT "The var genes of Plasmodium falciparum are located in the  
subtelomeric region of most chromosomes.";  
RL EMBO J. 15:4069-4077(1996).  
DR EMBL; U53324; G1297091; -;  
FT NON\_TER 1  
SQ SEQUENCE 2042 AA; 233669 MW; 77999777 CRC32;  
Query Match 71.9%; Score 46; DB 5; Length 2042;  
Best Local Similarity 100.0%; Pred. No. 1.22e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 658 PEDNPVE 664  
|:::|:  
QY 3 PEDNPVE 9  
RESULT 13  
ID Q92728 PRELIMINARY; PRT; 53 AA.  
AC Q92728;  
DT 01-FEB-1997 (TREMREL. 02, CREATED)  
DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE RETINOBLASTOMA SUSCEPTIBILITY PROTEIN (RB).  
GN RBL.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88320373.  
RA LEE E.Y., BOOKSTEIN R., YOUNG L.J., LIN C.J., ROSENFELD M.G.,  
RA LEE W.H.;  
RT "Molecular mechanism of retinoblastoma gene inactivation in  
retinoblastoma cell line Y79.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 85:6017-6021(1988).  
DR EMBL; M19701; G190988; -;  
SQ SEQUENCE 53 AA; 5392 MW; 2D5C3E39 CRC32;  
Query Match 70.3%; Score 45; DB 4; Length 53;  
Best Local Similarity 50.0%; Pred. No. 1.98e+01;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 27 PPPEEDPEQD 36  
|:::|:  
QY 1 APPDNPVED 10  
RESULT 14

```

ID P74580 PRELIMINARY; PRT; 86 AA.
AC P74580;
DT 01-FEB-1997 (TREMREL. 02, CREATED)
DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 9.9 KD PROTEIN.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA TABATA S.;
RN SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. PCC6803. II. Sequence determination of the entire
RT genome and assignment of potential protein-coding regions.";
RL DNA RES. 3:109-136(1996).
DR EMBL; D90916; D1019420; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 86 AA; 9909 MW; B95C9013 CRC32;

Query Match 70.3%; Score 45; DB 2; Length 86;
Best Local Similarity 62.5%; Pred. No. 1.98e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 49 PDDPVEE 56
QY 3 PEDNVED 10

RESULT 15
ID Q26168 PRELIMINARY; PRT; 97 AA.
AC Q26168;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE DNA FRAGMENT WITH HIGHLY REPEATED SEQUENCE (FRAGMENT).
OS PLASMIDIUM VIVAX.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA CAMPBELL J.R., FRANK E.D.;
RL SUBMITTED (APR-1989) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X15129; G10087; -.
FT NON_TER 1
SQ SEQUENCE 97 AA; 9736 MW; 3866D3C5 CRC32;

Query Match 70.3%; Score 45; DB 5; Length 97;
Best Local Similarity 71.4%; Pred. No. 1.98e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 13 PPEENPI 19
QY 2 PPEDNPV 8

```

Search completed: Thu Oct 21 15:33:32 1999  
Job time : 41 secs.

\*\*\*\*\*

MPERCH\_PP protein - protein database search, using Smith-Waterman algorithm

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Thu Oct 21 15:34:15 1999; MasPar time 3.32 Seconds  
Tabular output not generated. 64.049 Million cell updates/sec

Title: >US-09-040-485-7  
Description: (1-10) from US09040485.pep  
Perfect Score: 66  
Sequence: 1 EEOQEVPPDT 10

Scoring table: PAM 150  
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseg35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 14.889; Variance 49.397; scale 0.301

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	68.2	181	30	Mouse p26 protein.	2.23e+02
2	45	68.2	396	39	Soil derived peptide	2.23e+02
3	45	68.2	3144	29	Human huntingtin prot	2.23e+02
4	45	68.2	3144	11	Protein encoded by Hu	2.23e+02
5	45	68.2	3144	22	Human huntingtin.	2.23e+02
6	45	68.2	3144	26	Previously undescribe	2.23e+02
7	44	66.7	181	30	Rat p26 protein seque	2.80e+02
8	44	66.7	759	30	Homo sapiens telomera	2.80e+02
9	44	66.7	759	30	Human telomerase p105	2.80e+02
10	43	65.2	140	1	Gamma-chicken atrial	3.52e+02
11	43	65.2	414	14	Human TGF-beta 2 prot	3.52e+02
12	43	65.2	414	16	Transforming growth f	3.52e+02
13	43	65.2	442	1	Sequence encoded by h	3.52e+02
14	43	65.2	442	1	Human TGF-Beta2-442 p	3.52e+02
15	43	65.2	442	16	Human transforming gr	3.52e+02
16	43	65.2	442	4	Sequence of human tra	3.52e+02

17	43	65.2	700	1	R03663	Human myb related gen	3.52e+02
18	43	65.2	1757	39	W84351	Murine ubiquitin-prot	3.52e+02
19	43	65.2	1763	39	W83431	Rat Munc13-1.	3.52e+02
20	42	63.6	193	25	W23619	Prolactin antagonist	4.42e+02
21	42	63.6	225	1	P82078	Recombinant rat prepr	4.42e+02
22	42	63.6	226	3	R14599	Rat prolactin.	4.42e+02
23	42	63.6	392	2	R12345	Toxoplasma gondii pro	4.42e+02
24	42	63.6	428	2	R12352	Toxoplasma gondii P66	4.42e+02
25	42	63.6	473	15	R84637	Shc protein.	4.42e+02
26	42	63.6	473	21	W14005	Human SHC protein.	4.42e+02
27	42	63.6	474	17	R97243	SHC phosphotyrosine b	4.42e+02
28	42	63.6	494	11	R56519	Protein kinase (HRR25	4.42e+02
29	42	63.6	494	6	R31888	Defective tyrosine ki	4.42e+02
30	42	63.6	494	15	R76615	Saccharomyces cerevis	4.42e+02
31	41	62.1	120	31	W37933	Amino acid sequence o	5.53e+02
32	41	62.1	377	39	W85956	Triticum sp. cysteine	5.53e+02
33	41	62.1	508	32	W62558	Fibroblast growth fac	5.53e+02
34	41	62.1	570	26	W36002	Human Fchd531 gene pr	5.53e+02
35	41	62.1	571	37	W74889	Human secreted protei	5.53e+02
36	41	62.1	1345	30	W31346	Rat tumour suppressor	5.53e+02
37	41	62.1	1596	30	W31347	Rat tumour suppressor	5.53e+02
38	41	62.1	1719	16	R92100	Human R12 allele D283	5.53e+02
39	41	62.1	3119	29	W44743	Mouse huntingtin prot	5.53e+02
40	41	62.1	3119	26	W36888	Mouse huntingtin's di	5.53e+02
41	40	60.6	101	18	R90022	Ubi7 ubiquitin-lytic	6.91e+02
42	40	60.6	240	4	R22537	Foetal oncogene fem s	6.91e+02
43	40	60.6	1350	17	R91296	Drosophila nitric oxl	6.91e+02
44	40	60.6	1618	5	R27205	Human nestin.	6.91e+02
45	40	60.6	2115	32	W59276	Rubella virus RA27/3	6.91e+02

ALIGNMENTS

RESULT 1  
ID W44835 standard; peptide; 181 AA.  
AC W44835;  
DT 21-JUL-1998 \*(first entry)  
DE Mouse p26 protein.  
KW Mouse; p26; brain; hybridisation; dephosphorylase inhibitory activity;  
KW probe; haematogenesis.  
OS Mus sp.  
PN J09299092-A.  
PD 25-NOV-1997.  
PF 26-DEC-1996; 347877.  
PR 12-MAR-1996; JP-055196.  
PA (TAKE ) TAKEDA CHEM IND LTD.  
DR WPI: 98-056555/06.  
DR N-PSDB; V19305.  
PT Mammalian p26 proteins and their related DNA - useful for screening  
PT for de-phosphorylase inhibitory compounds  
PS Claim 1; Fig 3; 40pp; Japanese.  
CC This amino acid sequence represents the mouse p26 protein. The encoding  
CC gene sequence was isolated from a mouse cDNA library by phage plaque  
CC hybridisation using the rat p26 cDNA sequence (V19303) as a probe.  
CC The screening isolated the 882 bp sequence shown here. p26 protein is  
CC useful as a reagent for screening for compounds having dephosphorylase  
CC inhibitory activity. It is also useful as a treating and preventive  
CC agent for diseases related to the haematogenic system.  
SQ Sequence 181 AA;  
Query Match 68.2%; Score 45; DB 30; Length 181;  
Best Local Similarity 60.0%; Pred. No. 2.23e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 159 eeeeeppda 168  
|||:| |

QY 1 EEOQEVPPDT 10

RESULT 2

ID W82713 standard; Protein; 396 AA.

AC W82713;

DT 15-MAR-1999 (first entry)

DE Soil derived peptide synthase clone ps24 protein.  
 KW Peptide synthase; soil; lichen; antibiotic biosynthesis; humus;  
 KW therapeutic; immunosuppressor; antitumour agent; pathogen;  
 KW genetic diversity.  
 OS Unknown.  
 PN WO9853097-A2.  
 PD 26-NOV-1998.  
 PF 21-MAY-1998; CA0488.  
 PR 22-MAY-1997; US-861774.  
 PA (TERR-) TERRAGEN DIVERSITY INC.  
 PI Miao VFW, Seow KT, Waters B, Yap WH;  
 DR WPI; 99-070158/06.  
 DR N-PSDB; V82713.  
 PT New degenerate primers - used for recovering antibiotic biosynthetic  
 PT DNA from soil/lichen material  
 PS Claim 19; Page 87-89; 98pp; English.  
 CC This sequence represents a peptide synthase clone, isolated from soil.  
 CC This protein is used in a method for the recovery of antibiotic  
 CC biosynthetic DNA from humic materials or lichen. The PCR products of the  
 CC invention have the potential to be used as therapeutic molecules  
 CC including antibiotics, immunosuppressors and antitumour agents. The  
 CC method allows access to the reservoir of genetic diversity in soil  
 CC pathogenic micro-organisms, in order to find new antibiotics. It also  
 CC allows access to novel biosynthetic genes/enzymes that can be used to  
 CC produce antibiotics or produce specific compounds, enzymatically,  
 CC in vitro. 396 AA;  
 SQ Sequence 396 AA;

Query Match 68.2%; Score 45; DB 39; Length 396;  
 Best Local Similarity 60.0%; Pred. No. 2.23e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 355 gggqhvppqt 364  
 : || |||||  
 Qy 1 EEQEVPPDPT 10

RESULT 3  
 ID W44742 standard; Protein; 3144 AA.  
 AC W44742;  
 DT 01-JUN-1998 (first entry)  
 DE Human huntingtin protein.  
 KW Human; huntingtin gene; Huntington's disease; chromosome; marker;  
 KW locus; antisense; gene therapy; diagnosis.  
 OS Homo sapiens.  
 PN US5693757-A.  
 PD 02-DEC-1997.  
 PF 30-MAY-1995; 453265.  
 PR 20-MAY-1994; US-246982.  
 PR 05-MAR-1993; US-027498.  
 PR 01-JUL-1993; US-085000.  
 PR 30-MAY-1995; US-453265.  
 PA (GEO) GEN HOSPITAL CORP.  
 PI Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;  
 DR WPI; 98-031815/03.  
 DR N-PSDB; V05828.  
 PT Huntingtin protein and related nucleic acid - for diagnosis or  
 PT therapy of Huntington's disease  
 PS Claim 2; Fig 4; 112pp; English.  
 CC This is the amino acid sequence of the human huntingtin protein.  
 CC The gene sequence is characterised in that it contains a number of  
 CC CAG repeats in the 5' region (in this case 23 repeats). In healthy  
 CC individuals the usual number of CAG repeats in the gene is 11-34,  
 CC whereas in patients suffering from Huntington's disease (HD), the  
 CC number of repeats increases to 37-73 or 37-86. The huntingtin gene  
 CC spans 210 kb and encodes a protein of 348 kD. The gene is found in a  
 CC 500 kb region between the chromosomal markers D4S180 and D4S182 and is  
 CC preferentially mapped to the locus 4p16.3. The protein or the gene  
 CC encoding it, is useful for detecting a predisposition to develop HD,  
 CC for diagnosis and treatment of HD, especially by antisense and gene  
 CC therapy.  
 SQ Sequence 3144 AA;

Query Match 68.2%; Score 45; DB 29; Length 3144;  
 Best Local Similarity 60.0%; Pred. No. 2.23e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeeevdpt 2352  
 ||:|||||  
 Qy 1 EEQEVPPDPT 10

RESULT 4  
 ID R58777 standard; Protein; 3144 AA.  
 AC R58777;  
 DT 13-APR-1995 (first entry)  
 DE Protein encoded by Huntingtin DNA\IT15 gene.  
 KW Polymerase chain reaction; primer; PCR; amplify; Huntingtin; IT15;  
 KW open reading frame; polymorphic; (CAG)n; trinucleotide repeat;  
 KW allele; Huntington's disease; HD; chromosome; juvenile HD; unstable;  
 KW expandable; linkage disequilibrium.  
 OS Homo sapiens.  
 PN EP-614977-A.  
 PD 14-SEP-1994.  
 PF 07-MAR-1994; 301587.  
 PR 05-MAR-1993; US-027498.  
 PR 01-JUL-1993; US-085000.  
 PA (GEO) GEN HOSPITAL CORP.  
 PI Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;  
 DR WPI; 94-281205/35.  
 DR N-PSDB; Q67401.  
 PT New huntingtin protein and related nucleic acid, antibodies etc.  
 PT - for treatment and diagnosis of neuro-degenerative disease,  
 PT specifically Huntington's disease  
 PS Claim 1; Page 22-33; 66pp; English.  
 CC This sequence is encoded by a large gene, termed "Huntingtin" or  
 CC "IT15" which is present in the proximal part of the 0.5 mb segment  
 CC between members D4S180 and D4S182 on chromosome 4. The IT15 gene  
 CC spans about 210 kb and encodes a protein of approx. 348 kD. The  
 CC huntingtin open reading frame contains a polymorphic (CAG)n  
 CC trinucleotide repeat with at least 17 alleles in the normal  
 CC population, varying from about 11 to about 34 CAG copies. On  
 CC Huntington's disease (HD) chromosome, the length of the CAG repeat  
 CC is substantially increased, with between 37 to at least 73 copies.  
 CC This shows an apparent correlation with age of onset, the longest  
 CC segments are detected in juvenile HD cases. The presence of an  
 CC unstable, expandable trinucleotide repeat on HD chromosomes in the  
 CC region of the strongest linkage disequilibrium with the disorder  
 CC suggest that this alteration underlies the dominant phenotype of HD.  
 SQ Sequence 3144 AA;

Query Match 68.2%; Score 45; DB 11; Length 3144;  
 Best Local Similarity 60.0%; Pred. No. 2.23e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeeevdpt 2352  
 ||:|||||  
 Qy 1 EEQEVPPDPT 10

RESULT 5  
 ID W09871 standard; Protein; 3144 AA.  
 AC W09871;  
 DT 27-JUL-1997 (first entry)  
 DE Human huntingtin.  
 KW Huntingtin associated protein-1; HAP1; Huntington's disease.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT protein 1..230  
 FT /note= "Claim 20"  
 FN W09717443-A1.  
 PD 15-MAY-1997.  
 PR 08-NOV-1996; U17858.  
 PR 09-NOV-1995; US-556419.  
 PA (UJJO) UNIV JOHNS HOPKINS.  
 PI Lanahan A, Li S, Li X, Ross CA, Sharp AH, Snyder S;

PI Worley PF;  
 DR WPI: 97-281032/25.  
 PT Determination of the binding of huntingtin to huntingtin-associated  
 PT protein-1 - useful for screening for drugs for treating or  
 PT preventing Huntington's disease  
 PS Claim 20; Page 22-32; 69pp; English.  
 CC Human huntingtin (Hn) polypeptide (W09871) is the product of the  
 CC Huntington's disease (HD) locus. Proteins which specifically bind  
 CC to Hn, such as human huntingtin associated protein-1 (HAP1) (see  
 CC also W09870), can be used in assays for screening drug candidates.  
 CC The binding between Hn and HAP1 is enhanced by an expanded  
 CC polyglutamine repeat in Hn, the length of which correlates with the  
 CC time of disease onset. HAP1, in contrast to Hn, is expressed  
 CC selectively in the brain, suggesting that it may contribute to the  
 CC brain-specific pathology of HD. Hn, or portions of it, esp. amino  
 CC acids 1-230, or yeast calls expressing Hn, can be used to identify  
 CC cpds. that bind to, displace or prevent binding of Hn and HAP1.  
 CC The method is useful for screening candidate drugs for treating,  
 CC delaying onset of, or preventing HD.  
 SQ Sequence 3144 AA;

Query Match 68.2%; Score 45; DB 22; Length 3144;  
 Best Local Similarity 60.0%; Pred. No. 2.23e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeevdpnt 2352  
 II:::II I:I  
 QY 1 EEQEQVPPPT 10

RESULT 6  
 ID W36887 standard; Protein: 3144 AA.  
 AC W36887;  
 DT 13-MAR-1998 (first entry)  
 DE Previously undescribed protein encoded by a novel huntingtin (IT15) gene.  
 KW Huntingtin gene; IT15 gene; Huntington's disease; trinucleotide repeat;  
 KW neurodegenerative disorder; HD; gene therapy.  
 OS Homo sapiens.  
 PN US5686288-A.  
 PD 11-NOV-1997.  
 PF 20-MAY-1994; 246982.  
 PR 05-MAR-1993; US-027498.  
 PR 01-JUL-1993; US-085000.  
 PA (GEO) GEN HOSPITAL CORP.  
 PI Ambrose CM, Duyao WP, Gusella JF, MacDonald ME;  
 DR WPI: 97-558144/51.  
 DR N-PSDB; T97924.  
 PT Nucleic acid encoding huntingtin protein - useful for gene therapy  
 PT of Huntington's disease  
 PS Claim 2; Fig 4; 112pp; English.

CC The present sequence represents a previously undescribed protein,  
 CC encoded by a novel gene, termed huntingtin or IT15. The huntingtin  
 CC reading frame contains a polymorphic (CAG)<sub>n</sub> trinucleotide repeat with at  
 CC least 17 alleles in the normal population, varying from about 11 to 34  
 CC CAG copies. Huntington's disease (HD) is a progressive neurodegenerative  
 CC disorder characterised by motor disturbance, cognitive loss and  
 CC psychiatric manifestations. The genetic defect causing HD is assigned to  
 CC chromosome 4. On HD chromosomes, the length of the trinucleotide CAG  
 CC repeat is substantially increased, e.g. about 37 to at least 73 copies.  
 CC The huntingtin gene and proteins encoded by it, may be used for the  
 CC diagnosis or treatment of Huntington's disease. The huntingtin gene  
 CC is especially used in gene therapy of a symptomatic or presymptomatic  
 CC patient. The method comprises providing a functional huntingtin gene with  
 CC a (CAG)<sub>n</sub> repeat of the normal range of 11-34 copies, or an antisense  
 CC sequence, to the desired cells of the patient, in a manner that permits  
 CC the expression of the huntingtin protein provided by the gene, or  
 CC inhibits expression of the mutated huntingtin gene, for a time and in a  
 CC quantity sufficient to provide the huntingtin function to the cells of  
 CC the patient.  
 SQ Sequence 3144 AA;

Query Match 68.2%; Score 45; DB 26; Length 3144;  
 Best Local Similarity 60.0%; Pred. No. 2.23e+02;

Matches: 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 2343 eeeevdpnt 2352  
 II:::II I:I  
 QY 1 EEQEQVPPPT 10

RESULT 7  
 ID W4833 standard; peptide; 181 AA.  
 AC W4833;  
 DT 21-JUL-1998 (first entry)  
 DE Rat p26 protein sequence.  
 KW Rat; p26; brain; hybridisation; probe; primer extension; RACE;  
 KW amplification; dephosphorylase inhibitory activity; haematogenesis.  
 OS Rattus sp.  
 PN J09299092-A.  
 PD 25-NOV-1997.  
 PF 26-DEC-1996; 347877.  
 PR 12-MAR-1996; JP-055196.  
 PA (TAKE) TAKEDA CHEM IND LTD.  
 DR WPI: 98-056555/06.  
 DR N-PSDB; V19303.  
 PT Mammalian p26 proteins and their related DNA - useful for screening  
 PT for de-phosphorylase inhibitory compounds  
 PS Claim 1; Fig 1; 40pp; Japanese.  
 CC This amino acid sequence represents the rat p26 protein. The encoding  
 CC gene sequence was isolated from a rat brain cDNA library by phage plaque  
 CC hybridisation using the probes V19206 and V19207, followed by primer  
 CC extension. The screening isolated a product of 545 bp. The 5' and 3'  
 CC ends of the coding sequence were then obtained by RACE amplification  
 CC resulting in the isolation of the 875 bp rat p26 cDNA clone.  
 CC p26 protein is useful as a reagent for screening for compounds having  
 CC dephosphorylase inhibitory activity. It is also useful as a treating  
 CC and preventive agent for diseases related to the haematogenic system.  
 SQ Sequence 181 AA;

Query Match 66.7%; Score 44; DB 30; Length 181;  
 Best Local Similarity 60.0%; Pred. No. 2.80e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 159 eeeevdpnt 168  
 II:::II I:I  
 QY 1 EEQEQVPPPT 10

RESULT 8  
 ID W41927 standard; Protein: 759 AA.  
 AC W41927;  
 DT 21-JUL-1998 (first entry)  
 DE Homo sapiens telomerase protein p105.  
 KW telomerase; p105; treatment; prevention; cancer; restenosis;  
 KW inflammation; myocardial infarction; glomerulonephritis; transplant;  
 KW rejection; infection; HIV; human immunodeficiency virus;  
 KW bone marrow transplants; proliferation-restricted cells.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Domain 5..81  
 FT /note= "telomerase binding domain"  
 FT Domain 115..192  
 FT /note= "telomerase binding domain"  
 FT Domain 336..420  
 FT /note= "telomerase binding domain"  
 FT Domain 487..578  
 FT /note= "telomerase binding domain"  
 PN W09801543-A1.  
 PD 15-JAN-1998.  
 PF 08-JUL-1997; U12297.  
 PR 08-JUL-1996; US-676967.  
 PA (TULA-) TULARIK INC.  
 DR WPI: 98-101044/09.  
 DR N-PSDB; V13832.  
 PT New nucleic acid encoding human telomerase protein p105 or its

PT fragments - used for therapeutic modulation of telomerase activity  
 PT and for screening for potential modulators of telomerase-target  
 PT binding  
 PS Claim 1; Pages 17-19; 32pp; English.  
 CC The sequence is that of telomerase protein p105, it can be used to  
 CC screen for agents, e.g. antibodies, that modulate binding of human  
 CC telomerase to its binding target. Those that inhibit telomerase  
 CC activity can be used to treat conditions such as cancer, restenosis,  
 CC inflammation, myocardial infarction, glomerulonephritis, transplant  
 CC rejection and infections (e.g. with human immunodeficiency virus),  
 CC while those that are agonists can be used to extend the life of  
 CC proliferation-restricted cells, especially normal somatic cells,  
 CC e.g. in cases of hypersensitivity or atrophy, also to improve  
 CC production of recombinant proteins by maximising cell density and  
 CC survival and expansion of precursor cells being used for bone marrow  
 CC transplants. They may also be used for diagnosis. Other uses of  
 CC telomerase proteins are isolation, enrichment and concentration  
 CC of telomerase RNA or proteins; as immunogens; in therapy; as reagent  
 CC where nascent oligonucleotides of known structure are needed (e.g.  
 CC for tagging native nucleic acid molecules) and for regulating cell  
 CC growth/density tolerance. The agents and the telomerase proteins  
 CC should be very specific, e.g. they are selective for cancer cells  
 CC without harming somatic cells.  
 SQ Sequence 759 AA;

Query Match 66.7%; Score 44; DB 30; Length 759;  
 Best Local Similarity 66.7%; Pred. No. 2.80e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 629 eeqskvppe 637  
 ||| |||  
 Qy 1 EEQEVPPD 9

RESULT 9  
 ID W46593 standard; Protein; 759 AA.  
 AC W46593;  
 DT 06-JUL-1998 (first entry)  
 DE Human telomerase p105 subunit.  
 KW Telomerase; p105; human; cell replication; cancer; restenosis;  
 KW multiple sclerosis; inflammation; rheumatoid arthritis;  
 KW myocardial infarction; glomerulonephritis; transplant rejection;  
 KW infection; therapy.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Domain 5..81 /note= "RRM1 binding domain"  
 FT Domain 115..192 /note= "RRM2 binding domain"  
 FT Domain 336..420 /note= "RRM3 binding domain"  
 FT Domain 487..578 /note= "RRM4 binding domain"  
 PN W09801542-A1.  
 PD 15-JAN-1998.  
 PF 08-JUL-1997; U12296.  
 PR 08-JUL-1996; US-676974.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Collins K;  
 DR WPI: 98-101043/09.  
 DR N-PSDB: V05369-72.  
 PT New nucleic acid encoding human telomerase proteins or their  
 PT fragments - useful for therapeutic modulation of telomerase activity  
 PT and for screening for potential modulators of telomerase-target  
 PT binding  
 PS Claim 4; Page 17-19; 32pp; English.  
 CC This protein comprises the p105 subunit of human telomerase. p105  
 CC can be isolated from human cells or expressed in host cells using  
 CC native p105 cDNA (see V05369), or optimised synthetic sequences  
 CC (see V05370-72). The invention provides methods relating to human  
 CC telomerase and related nucleic acids, including the subunit  
 CC proteins p140, p105, p48 and p43. The invention also provides  
 CC isolated telomerase hybridisation probes and primers capable of

CC specifically hybridising with the telomerase gene, telomerase-  
 CC specific binding agents such as specific antibodies, and methods  
 CC of making and using the subject compositions in diagnosis (e.g.  
 CC genetic hybridisation screens for telomerase transcripts), therapy  
 CC (e.g. gene therapy to modulate telomerase gene expression) and in  
 CC the biopharmaceutical industry (e.g. reagents for screening  
 CC chemical libraries for lead agents). Modulation of telomerase  
 CC expression can be used for the treatment or prevention of cancer,  
 CC restenosis, inflammation, myocardial infarction, glomerulonephritis,  
 CC transplant rejection or infections (e.g. with HIV). Telomerase  
 CC proteins can also be used in the isolation, enrichment and  
 CC concentration of telomerase RNA proteins, as immunogens, in  
 CC therapy, for regulating cell growth/density tolerance and for  
 CC polymerising nucleic acid on a substrate.  
 SQ Sequence 759 AA;

Query Match 66.7%; Score 44; DB 30; Length 759;  
 Best Local Similarity 66.7%; Pred. No. 2.80e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 629 eeqskvppe 637  
 ||| |||  
 Qy 1 EEQEVPPD 9

RESULT 10  
 ID R05669 standard; peptide; 140 AA.  
 AC R05669;  
 DT 15-AUG-1990 (first entry)  
 DE Gamma-chicken atrial natriuretic peptide.  
 KW Gamma-chicken atrial natriuretic peptide; diuretic; hypertensive.  
 OS Gallus sp.  
 FH Key Location/Qualifiers  
 FT disulfide\_bond 118..134  
 FT region 25..140 /label=Gamma-chAMP specific.  
 FT J02025499-A.  
 PD 26-JAN-1990.  
 PR 14-JUL-1988; 173739.  
 PR 14-JUL-1988; JP-173739.  
 PA (MATS/) Matsuo T.  
 DR WPI: 90-071804/10.  
 DR N-PSDB: Q03465.  
 PT Physio-active peptide derived from birds - has cysteine bridge,  
 PT specified amino acid sequence and diuretic and hypertensive properties.  
 PS Disclosure; Fig 13; 16pp; Japanese.  
 CC Gamma-chAMP is obtained from the 12 kD fraction obtained from treated  
 CC homogenised chicken heart tissue.  
 CC See also Q03466-Q03468, R03301 and R03302.  
 SQ Sequence 140 AA;

Query Match 65.2%; Score 43; DB 1; Length 140;  
 Best Local Similarity 44.4%; Pred. No. 3.52e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 66 qtqeippe 74  
 : |||:  
 Qy 1 EEQEVPPD 9

RESULT 11  
 ID R73597 standard; Protein; 414 AA.  
 AC R73597;  
 DT 20-DEC-1995 (first entry)  
 DE Human TGF-beta 2 protein.  
 KW Transforming growth factor-beta; Human TGF-beta protein; TGF-beta 1;  
 KW TGF-beta 2; TGF-beta 3; osteogenic cell source; OCS; bone deficiency;  
 KW bone-inducing cofactor.  
 OS Homo sapiens.  
 PN US5409896-A.  
 PD 25-APR-1995.  
 PF 12-NOV-1993; 401906.  
 PR 01-SEP-1989; US-401906.



PR 12-NOV-1991; US-790856.  
 PR 18-MAY-1993; US-063841.  
 PR 12-NOV-1993; US-132405.  
 PA (GETH ) GENENTECH INC.  
 PI Ammann AJ, Rudman CG;  
 DR WPI: 95-169610/22.  
 PT Compens. for treating skeletal tissue deficiency - comprising  
 PT transforming growth factor-beta and an osteogenic cell source in a  
 PT carrier  
 PS Disclosure: Column 17-20: 19pp; English.  
 CC This sequence represents human transforming growth factor-beta 2  
 CC (TGF-beta 2). The sequences for human TGF-beta 1 (see R73596) and human  
 CC TGF-beta 3 (see R73598) are claimed within the scope of the invention.  
 CC The invention is a composition consisting of a TGF-beta protein and an  
 CC osteogenic cell source (OCS) formulated in an acceptable carrier other  
 CC than a bone morphogenic cofactor. This composition can be used for the  
 CC restoration of bone deficiency. This provides for the generation of  
 CC mature bone only where it is required, without the inclusion of a  
 CC specific bone-inducing cofactor. This method can be used with any of the  
 CC 5 human TGF-beta's or with TGF-beta from other species.  
 SQ Sequence 414 AA;

Query Match 65.28; Score 43; DB 14; Length 414;  
 Best Local Similarity 62.58; Pred. No. 3.52e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66  
 QY 2 EQQEVPPD 9

RESULT 12  
 ID R83055 standard; Protein; 414 AA.  
 AC R83055;  
 DT 25-JUN-1996 (first entry)  
 DE Transforming growth factor-beta 2.  
 KW macrophage inducible nitric oxide synthase; iNOS; constitutive NOS;  
 KW interleukin-1-beta; transforming growth factor-beta; TGF-beta; IL1-beta;  
 KW nitric oxide production; hypotension; inflammation; septic shock;  
 KW treatment.  
 OS Mammalian sp.  
 FH Key Location/Qualifiers  
 FT protein 303..414  
 FT /note= "represents the mature active TGF beta-1 mol."  
 PN W09526745-A1.  
 PD 12-OCT-1995.  
 PF 05-APR-1994; U03705.  
 PR 05-APR-1994; WO-U03705.  
 PA (HARD ) HARVARD COLLEGE.  
 PI Lee M, Perrella MA;  
 DR WPI: 95-358443/46.  
 DR N-PSDB; T05877.  
 PT Treatment of hypotension, esp. in septic shock - by administering  
 PT transforming growth factor-beta e.g. to inhibit inducible nitric  
 PT oxide synthase gene transcription  
 PS Disclosure: Fig 18: 52pp; English.  
 CC Transforming growth factor-beta 2 (TGF-beta 2) has been found to inhibit  
 CC inducible nitric oxide synthase (iNOS) gene transcription, esp. in  
 CC interleukin-1-beta (IL1-beta) stimulated rat smooth muscle cells, and at  
 CC a dose which does not inhibit constitutive NOS. TGF-beta 1 (R83054) or 2  
 CC or their active fragments (esp. derived from the carboxy-terminal 112  
 CC amino acids), can be used in the treatment of hypotension, such as that  
 CC associated with severe inflammation or septic shock.  
 SQ Sequence 414 AA;

Query Match 65.28; Score 43; DB 16; Length 414;  
 Best Local Similarity 62.58; Pred. No. 3.52e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66  
 QY 2 EQQEVPPD 9

RESULT 13  
 ID P91899 standard; protein; 442 AA.  
 AC P91899;  
 DT 23-DEC-1990 (first entry)  
 DE Sequence encoded by human transforming growth factor (TGF) beta-2  
 DE precursor 442 cDNA in pPC-21  
 KW Cell differentiation; cell proliferation.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT region 20..442  
 FT /note="Claimed"  
 FT peptide 4..19  
 FT /note="Signal"  
 FT cleavage\_site 20..21  
 FT region 110..144  
 FT /note="This entire SQ is replaced with Asn in simian  
 FT TGF-beta-2-414"  
 FT modified\_site 72  
 FT /note="Potential glycosylation site"  
 FT modified\_site 168  
 FT /note="Potential glycosylation site"  
 FT modified\_site 269  
 FT /note="Potential glycosylation site"  
 FT modified\_site 331..442  
 FT protein  
 FT D53833897-A.  
 PD 03-MAY-1989.  
 PF 05-OCT-1988; 833897.  
 PR 18-AUG-1988; US-234065,  
 PA (ONCO-) Oncogen.  
 PI Purchio AF, Madisen L, Webb N;  
 DR WPI: 89-138796/19.  
 DR N-PSDB; N90767.  
 PT New DNA sequence encoding transforming growth factor beta 2 -  
 PT used for large scale expression in eucaryotic cells  
 PS Claim 4; Fig 1a; 27pp; German.  
 CC Polya-RNA was isolated from the tamoxifen-treated, human prostatic  
 CC adenocarcinoma line PC-3 and converted to cDNA. TGF DNA is pref. used  
 CC for control of the SV40 promoter, and expressed in CHO cells. The simian  
 CC SQ is also claimed.  
 SQ Sequence 442 AA;

Query Match 65.28; Score 43; DB 1; Length 442;  
 Best Local Similarity 62.58; Pred. No. 3.52e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66  
 QY 2 EQQEVPPD 9

RESULT 14  
 ID R05748 standard; protein; 442 AA.  
 AC R05748;  
 DT 02-NOV-1990 (first entry)  
 DE Human TGF-Beta2-442 precursor.  
 KW Human TGF-Beta2 precursor; cancer; tumoricide; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT protein 331..442  
 FT peptide 4..19  
 FT /label=Sinal peptide.  
 PN EP-376785-A.  
 PD 4-JUL-1990.  
 PF 14-DEC-1989; 403480.  
 PR 16-DEC-1988; US-285140.  
 PR 5-DEC-1989; US-446020.  
 PA (ONCO-) Oncogen Ltd Partner.  
 PI Purchio AF, Madisen L, Webb N;  
 DR WPI: 90-203127/27.  
 DR N-PSDB; Q05126  
 PT Cloning and expression of transforming growth factor beta 2 -  
 PT used for treatment of tumors or for augmenting wound healing.

PS Claim 1; Fig 1a; 58pp; English.  
CC TGF-beta2 may be used in treatment of tumors at effective doses,  
CC and may also be useful in augmenting wound healing by stimulating  
CC cell proliferation. The growth factor can be produced at high  
CC levels from a CHO expression system.  
SQ Sequence 442 AA;

Query Match 65.2%; Score 43; DB 1; Length 442;  
Best Local Similarity 62.5%; Pred. No. 3.52e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66  
| :|||:  
QY 2 EQQEVPPD 9

RESULT 15  
ID R79922 standard; Protein; 442 AA.  
AC R79922;  
DT 28-MAY-1996 (first entry)  
DE Human transforming growth factor-2.  
KW TGF-beta1; TGF-beta2; transforming growth factor; protein;  
KW cell differentiation; cell proliferation; CHO; Chinese hamster;  
KW ovary; COS; monkey kidney; animal; mammal.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT peptide 4..19  
FT /note= "signal peptide"  
FT cleavage\_site 20..21  
FT /note= "putative signal sequence cleavage site"  
FT peptide 331..442  
FT /note= "mature peptide"  
PN EP-676474-A1.  
PD 11-OCT-1995.  
PF 14-DEC-1989; 104223.  
PR 16-DEC-1988; US-285140.  
PR 05-DEC-1989; US-446020.  
PA (ONCO ) ONCOGEN LP.  
PI Madisen L, Purchio AF, Webb N;  
DR WPI; 95-346094/45.  
DR N-PSDB; T04116.  
PT Hybrid transforming growth factor beta-1/TGF-beta-2 precursor - used  
PT to produce biologically active, mature TGF-beta-2  
PS Disclosure; Fig.1a; 58pp; English.  
CC This sequence is expressed in a host cell, preferably a  
CC COS or CHO cell, so the host cell produces active TGF-beta2. The  
CC produced TGF-beta2 protein can be used to regulate cellular  
CC differentiation and proliferation.  
SQ Sequence 442 AA;

Query Match 65.2%; Score 43; DB 16; Length 442;  
Best Local Similarity 62.5%; Pred. No. 3.52e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66  
| :|||:  
QY 2 EQQEVPPD 9

Search completed: Thu Oct 21 15:34:36 1999  
Job time : 21 secs.

\*\*\*\*\*  
M P S R E L H  
\*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:39:36 1999; MasPar time 1.51 Seconds  
Tabular output not generated. 77.591 Million cell updates/sec

Title: >US-09-040-485-7  
Description: (1-10) from US09040485.pep  
Perfect Score: 66  
Sequence: 1 EEQEEVPPDT 10

Scoring table: PAM 150  
Gap 15

Searched: 119857 seqs, 11713122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 14.000; Variance 47.171; scale 0.297

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	68.2	3144	2	US-08-457- Sequence 42, Applicati	1.15e+02
2	45	68.2	3144	1	US-08-453- Sequence 6, Applicatio	1.15e+02
3	45	68.2	3144	1	US-08-246- Sequence 6, Applicatio	1.15e+02
4	44	66.7	759	2	US-09-098- Sequence 1, Applicatio	1.45e+02
5	44	66.7	759	1	US-08-676- Sequence 1, Applicatio	1.45e+02
6	44	66.7	759	2	US-08-676- Sequence 1, Applicatio	1.45e+02
7	43	65.2	16	4	5221620-9 Patent No. 5221620.	1.83e+02
8	43	65.2	414	1	US-08-395- Sequence 2, Applicatio	1.83e+02
9	43	65.2	414	1	US-08-132- Sequence 2, Applicatio	1.83e+02
10	43	65.2	414	3	PCT-US91-0 Patent No. 5221620.	1.83e+02
11	43	65.2	414	3	PCT-US94-0 Patent No. 5221620.	1.83e+02
12	43	65.2	414	3	PCT-US94-0 Patent No. 5221620.	1.83e+02
13	43	65.2	414	3	PCT-US94-0 Patent No. 5221620.	1.83e+02
14	42	63.6	423	1	US-08-353- Sequence 3, Applicatio	2.30e+02
15	42	63.6	423	2	US-08-551- Sequence 3, Applicatio	2.30e+02
16	42	63.6	423	2	US-08-363- Sequence 1, Applicatio	2.30e+02
17	42	63.6	423	2	US-08-363- Sequence 1, Applicatio	2.30e+02
18	42	63.6	423	2	US-08-363- Sequence 1, Applicatio	2.30e+02
19	42	63.6	474	1	US-08-353- Sequence 8, Applicatio	2.30e+02
20	42	63.6	474	1	US-08-551- Sequence 8, Applicatio	2.30e+02
21	42	63.6	494	1	US-08-454- Sequence 2, Applicatio	2.30e+02
22	42	63.6	494	1	US-08-453- Sequence 2, Applicatio	2.30e+02
23	42	63.6	494	1	US-08-447- Sequence 2, Applicatio	2.30e+02

24	41	62.1	120	2	US-08-727- Sequence 2, Applicatio	2.89e+02
25	41	62.1	120	1	US-08-253- Sequence 37, Applicati	2.89e+02
26	41	62.1	559	2	US-08-884- Sequence 6, Applicatio	2.89e+02
27	41	62.1	1346	2	US-08-635- Sequence 2, Applicatio	2.89e+02
28	41	62.1	1719	2	US-08-399- Sequence 4, Applicatio	2.89e+02
29	41	62.1	1719	2	US-08-459- Sequence 8, Applicatio	2.89e+02
30	41	62.1	3118	2	US-08-457- Sequence 8, Applicatio	2.89e+02
31	41	62.1	3119	1	US-08-453- Sequence 16, Applicati	2.89e+02
32	41	62.1	3119	1	US-08-246- Sequence 16, Applicati	2.89e+02
33	40	60.6	368	1	US-08-176- Sequence 15, Applicati	3.62e+02
34	40	60.6	368	2	US-08-461- Sequence 15, Applicati	3.62e+02
35	40	60.6	412	2	US-08-741- Sequence 2, Applicatio	3.62e+02
36	40	60.6	706	2	US-08-007- Sequence 4, Applicatio	3.62e+02
37	40	60.6	706	1	US-08-339- Sequence 29, Applicati	3.62e+02
38	40	60.6	706	2	US-08-689- Sequence 4, Applicatio	3.62e+02
39	40	60.6	763	1	US-08-424- Sequence 13, Applicati	3.62e+02
40	40	60.6	763	1	US-08-155- Sequence 13, Applicati	3.62e+02
41	40	60.6	763	3	PCT-US93-1 Sequence 13, Applicati	3.62e+02
42	40	60.6	1350	2	US-08-424- Sequence 13, Applicati	3.62e+02
43	40	60.6	1350	2	US-08-319- Sequence 9, Applicatio	3.62e+02
44	40	60.6	1618	1	US-07-853- Sequence 4, Applicatio	3.62e+02
45	40	60.6	2205	1	US-08-093- Sequence 2, Applicatio	3.62e+02

ALIGNMENTS

RESULT 1  
ID US-08-457-273B-42 STANDARD; PRT; 3144 AA.  
XX  
AC xxxxxx  
XX  
DT

Sequence 42, Application US/08457273B

Sequence 42, Application US/08457273B

Patent No. 5849995

GENERAL INFORMATION:

APPLICANT: Hayden, Michael

APPLICANT: Lin, Biaoyang

APPLICANT: Nasir, Jamal

TITLE OF INVENTION: Mouse Model for Huntington's Disease and

TITLE OF INVENTION: Related DNA Sequences

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESS: Virginia Bennett

STREET: PO Box 37428

CITY: Raleigh

STATE: No. 5849995th Carolina

COUNTRY: US

ZIP: 27627

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/457,273B

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Bennett, Virginia C.

REGISTRATION NUMBER: 37,092

REFERENCE/DOCKET NUMBER: 3477-85A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-854-1400

TELEFAX: 919-854-1401

INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENGTH: 3144 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

```

AC      xxxxxx
AD      XX
AE      DT
AF      XX
AG      XX
AH      DE
AI      XX
AJ      XX
AK      Sequence 6, Application US/08246982A
AL      Sequence 6, Application US/08246982A
AM      Patent No. 5686288
AN      GENERAL INFORMATION:
AO      APPLICANT: MacDonald, Marcy E.
AP      APPLICANT: Ambrose, Christine M.
AQ      APPLICANT: Duyao, Mabel P.
AR      APPLICANT: Gusella, James F.
AS      TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
AT      NUMBER OF SEQUENCES: 25
AU      CORRESPONDENCE ADDRESS:
AV      ADDRESSEE: Sterne, Kessler, Goldstein & Fox
AW      STREET: 1100 New York Avenue
AX      CITY: Washington
AY      STATE: D.C.
AZ      COUNTRY: U.S.A.
BA      ZIP: 20005
BB      COMPUTER READABLE FORM:
BC      MEDIUM TYPE: Floppy disk
BD      COMPUTER: IBM PC compatible
BE      OPERATING SYSTEM: PC-DOS/MS-DOS
BF      SOFTWARE: PatentIn Release #1.0, Version #1.25
BG      CURRENT APPLICATION DATA:
BH      APPLICATION NUMBER: US/08/246,982A
BI      FILING DATE: May 20, 1994
BJ      CLASSIFICATION: 435
BK      ATTORNEY/AGENT INFORMATION:
BL      NAME: Goldstein, Jorge, A.
BM      REGISTRATION NUMBER: 29,021
BN      REFERENCE/DOCKET NUMBER: 0609.3880002
BO      TELECOMMUNICATION INFORMATION:
BP      TELEPHONE: (202) 371-2600
BQ      TELEFAX: (202) 371-2540
BR      INFORMATION FOR SEQ ID NO: 6:
BS      SEQUENCE CHARACTERISTICS:
BT      LENGTH: 3144 amino acids
BU      TYPE: amino acid
BV      TOPOLOGY: linear
BW      MOLECULE TYPE: protein
BX      SEQUENCE 3144 AA; 347896 MW; 52186077 CN;
BY      Query Match 68.28; Score 45; DB 1; Length 3144;
BZ      Best Local Similarity 60.0%; Pred.No.1.15e+02;
CA      Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
CB
CC      Db 2343 EEEVDVDPNT 2352
CD      QY 1 EEQGEVPPDT 10
CE      RESULT 4
CF      ID US-09-098-487-1 STANDARD; PRT; 759 AA.
CG      XX xxxxxxx
CH      XX
CI      DT
CJ      XX
CK      XX
CL      Sequence 1, Application US/09098487
CM      Sequence 1, Application US/09098487
CN      Patent No. 5917025
CO      GENERAL INFORMATION:
CP      APPLICANT: COLLINS, Kathleen
CQ      TITLE OF INVENTION: Human Telomerase
CR      NUMBER OF SEQUENCES: 11
CS      CORRESPONDENCE ADDRESS:
CT      ADDRESSEE: Science & Technology Law Group
CU      STREET: 268 Bush Street, Suite 3200
CV

```

CC CITY: San Francisco  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94104  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/09/098,487  
CC FILING DATE:  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Osman Ph.D., Richard A  
CC REGISTRATION NUMBER: 36,627  
CC REFERENCE/DOCKET NUMBER: UCB96-055  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415)343-4341  
CC TELEFAX: (415)343-4342  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 759 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: not relevant  
CC MOLECULE TYPE: peptide  
CC SEQUENCE: 759 AA; 85737 MW; 2781856 CN;

Query Match 66.7%; Score 44; DB 2; Length 759;  
Best Local Similarity 66.7%; Pred. No. 1.45e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 629 EQSKVPPE 637  
QY 1 EEQEVPPD 9

RESULT 5  
ID US-08-676-967-1 STANDARD; PRT: 759 AA.

XX xxxxxx

Sequence 1, Application US/08676967

Sequence 1, Application US/08676967

Patent No. 5747317

GENERAL INFORMATION:

APPLICANT: COLLINS, KATHLEEN

TITLE OF INVENTION: Human Telomerase

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Science & Technology Law Group

STREET: 268 Bush Street, Suite 3200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/676,967

FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Osman Ph.D., Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: UCB96-055

CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415)343-4341  
CC TELEFAX: (415)343-4342  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 759 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: not relevant  
CC MOLECULE TYPE: peptide  
CC SEQUENCE: 759 AA; 85737 MW; 2781856 CN;

Query Match 66.7%; Score 44; DB 1; Length 759;  
Best Local Similarity 66.7%; Pred. No. 1.45e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 629 EQSKVPPE 637

QY 1 EEQEVPPD 9

RESULT 6  
ID US-08-676-974-1 STANDARD; PRT: 759 AA.

XX xxxxxx

Sequence 1, Application US/08676974

Sequence 1, Application US/08676974

Patent No. 5770422

GENERAL INFORMATION:

APPLICANT: COLLINS, KATHLEEN

TITLE OF INVENTION: Human Telomerase

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Science & Technology Law Group

STREET: 268 Bush Street, Suite 3200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/676,974

FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Osman Ph.D., Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: UCB96-055

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415)343-4341

TELEFAX: (415)343-4342

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 759 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: not relevant

MOLECULE TYPE: peptide

SEQUENCE: 759 AA; 85737 MW; 2781856 CN;

Query Match 66.7%; Score 44; DB 2; Length 759;  
Best Local Similarity 66.7%; Pred. No. 1.45e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 629 EQSKVPPE 637

QY 1 EEEVPPD 9

RESULT 7  
ID 5221620-9 STANDARD; PRT; 17 AA.  
XX AC xxxxxx  
XX 01-JAN-1900  
XX Patent No. 5221620.  
XX Patent No. 5221620  
XX APPLICANT: PURCHIO, ANTHONY F.; MADISEN, LINDA; WEBB, NANCY  
XX TITLE OF INVENTION: CLONING AND EXPRESSION OF TRANSFORMING  
XX GROWTH FACTOR BETA-2  
XX NUMBER OF SEQUENCES: 16  
XX CURRENT APPLICATION DATA:  
XX APPLICATION NUMBER: US/07/446,020  
XX FILING DATE: 05-DEC-1989  
XX PRIOR APPLICATION DATA:  
XX APPLICATION NUMBER: 285,140  
XX FILING DATE: 16-DEC-1988  
XX APPLICATION NUMBER: 234,065  
XX FILING DATE: 18-AUG-1988  
XX APPLICATION NUMBER: 148,267  
XX FILING DATE: 25-JAN-1988  
XX APPLICATION NUMBER: 106,752  
XX FILING DATE: 06-OCT-1987  
XX SEQ ID NO: 9:  
XX LENGTH: 16  
XX SEQUENCE 17 AA; 1922 MW; 1971 CN;

Query Match 65.2%; Score 43; DB 4; Length 16;  
Best Local Similarity 62.5%; Pred. No. 1.83e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 EPEVPPE 16  
QY 2 EEEVPPD 9

RESULT 8  
ID US-08-395-939A-2 STANDARD; PRT; 414 AA.  
XX AC xxxxxx

Sequence 2, Application US/08395939A  
Sequence 2, Application US/08395939A  
Patent No. 5604204  
GENERAL INFORMATION:  
APPLICANT: Ammann, Arthur J.  
APPLICANT: Rudman, Christopher G.  
TITLE OF INVENTION: TGF-BETA COMPOSITION FOR INDUCING BONE  
GROWTH  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/395,939A  
CC FILING DATE: 27-FEB-1995  
CC CLASSIFICATION: 514  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/132405  
CC FILING DATE: 12-NOV-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/063841  
CC FILING DATE: 18-MAY-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/790856  
CC FILING DATE: 12-NOV-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/401906  
CC FILING DATE: 1-SEP-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hasak, Janet E.  
CC REGISTRATION NUMBER: 28,616  
CC REFERENCE/DOCKET NUMBER: P0597D1C2D1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 415/225-1896  
CC TELEFAX: 415/952-9881  
CC TELEX: 910/371-7168  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 414 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC SEQUENCE 414 AA; 47747 MW; 890443 CN;

Query Match 65.2%; Score 43; DB 1; Length 414;  
Best Local Similarity 62.5%; Pred. No. 1.83e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 EPEVPPE 66  
QY 2 EEEVPPD 9

RESULT 9  
ID US-08-132-405-2 STANDARD; PRT; 414 AA.  
XX AC xxxxxx

Sequence 2, Application US/08132405  
Sequence 2, Application US/08132405  
Patent No. 5409896  
GENERAL INFORMATION:  
APPLICANT: Ammann, Arthur J.  
APPLICANT: Rudman, Christopher G.  
TITLE OF INVENTION: Method of Inducing Bone Growth Using  
TGF-Beta  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/132,405  
FILING DATE: 06-OCT-1993  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:



CC TITLE OF INVENTION: TRANSCRIPTION  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Fish & Richardson  
CC STREET: 225 Franklin Street  
CC CITY: Boston  
CC STATE: Massachusetts  
CC COUNTRY: U.S.A.  
CC ZIP: 02110-2804  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
CC COMPUTER: IBM PS/2 Model 502 or 55SX  
CC OPERATING SYSTEM: MS-DOS (Version 5.0)  
CC SOFTWARE: WordPerfect (Version 5.1)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US94/03705  
CC FILING DATE: 5 April 1994  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Janis K. Fraser  
CC REGISTRATION NUMBER: Reg. No. 34,819  
CC REFERENCE/DOCKET NUMBER: 05433/007001  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 542-5070  
CC TELEFAX: (617) 542-8906  
CC TELEX: 200154  
CC INFORMATION FOR SEQ ID NO: 6:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 414  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC SEQUENCE 414 AA; 47747 MW; 890443 CN;

Query Match 65.2%; Score 43; DB 3; Length 414;  
Best Local Similarity 62.5%; Pred. No. 1.83e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 EPEEVPPE 66  
1 :||||:  
Qy 2 EQQEVPPD 9

RESULT 13  
ID 5221620-2 STANDARD; PRT; 479 AA.  
XX  
AC xxxxxx  
XX  
DT 01-JAN-1900  
XX  
DE Patent No. 5221620.  
XX  
CC Patent No. 5221620  
CC APPLICANT: PURCHIO, ANTHONY F.; MADISEN, LINDA; WEBB, NANCY  
CC TITLE OF INVENTION: CLONING AND EXPRESSION OF TRANSFORMING  
CC GROWTH FACTOR BETA-2  
CC NUMBER OF SEQUENCES: 16  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/446,020  
CC FILING DATE: 05-DEC-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 285,140  
CC FILING DATE: 16-DEC-1988  
CC APPLICATION NUMBER: 234,065  
CC FILING DATE: 18-AUG-1988  
CC APPLICATION NUMBER: 148,267  
CC FILING DATE: 25-JAN-1988  
CC APPLICATION NUMBER: 106,752  
CC FILING DATE: 06-OCT-1987  
CC SEQ ID NO: 2

CC LENGTH: 442  
SQ SEQUENCE 479 AA; 54648 MW; 1314476 CN;

Query Match 65.2%; Score 43; DB 4; Length 442;  
Best Local Similarity 62.5%; Pred. No. 1.83e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 EPEEVPPE 66  
1 :||||:  
Qy 2 EQQEVPPD 9

RESULT 14  
ID US-08-353-550-3 STANDARD; PRT; 423 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
XX  
Sequence 3, Application US/08353550  
Sequence 3, Application US/08353550  
Patent No. 5744313  
GENERAL INFORMATION:  
APPLICANT: Williams, Lewis T.  
APPLICANT: Cavanaugh, William M.  
TITLE OF INVENTION: No. 5744313e1 Protein Domain Which Binds  
TITLE OF INVENTION: Tyrosine Phosphorylated Proteins  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: One Market Plaza, Steuart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,550  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Dow, Karen B.  
REGISTRATION NUMBER: 29,684  
REFERENCE/DOCKET NUMBER: 02307K-057300  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 423 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 423 AA; 45955 MW; 898238 CN;

Query Match 63.6%; Score 42; DB 1; Length 423;  
Best Local Similarity 55.6%; Pred. No. 2.30e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 218 DEEEVEPPD 226  
1 :|||:  
Qy 1 EQQEVPPD 9

RESULT 15  
ID US-08-551-687-3 STANDARD; PRT; 423 AA.  
XX



```

AC xxxxxx
XX
XX
DT
XX
XX
DE
XX
Sequence 3, Application US/08551687
Sequence 3, Application US/08551687
Patent No. 5925547
GENERAL INFORMATION:
CC APPLICANT: Williams, Lewis T.
CC APPLICANT: Cavanaugh, William M.
CC TITLE OF INVENTION: No. 5925547el Protein Domain Which Binds
CC TITLE OF INVENTION: Tyrosine Phosphorylated Proteins
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend and Crew
CC STREET: One Market Plaza, Steuart Tower, Suite 2000
CC City: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94105
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/551,687
CC FILING DATE: 01-NOV-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/353,550
CC FILING DATE: 09-DEC-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Murphy, Matthew B.
CC REGISTRATION NUMBER: P39,787
CC REFERENCE/DOCKET NUMBER: 2307K-5731
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-326-2400
CC TELEFAX: 415-326-2422
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 423 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 423 AA; 45955 MW; 898238 CN;

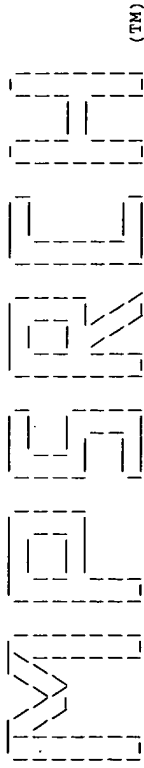
Query Match 63.6%; Score 42; DB 2; Length 423;
Best Local Similarity 55.8%; Pred. No. 2.30e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 218 DEEEEPD 226
QY 1 EEQVEVPD 9

```

Search completed: Thu Oct 21 15:39:43 1999  
Job time : 7 secs.

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:36:55 1999; MasPar time 3.05 Seconds  
Tabular output not generated. 131.513 Million cell updates/sec

Title: >US-09-040-485-7  
Description: (1-10) from US09040485.pep  
Perfect score: 66  
Sequence: 1 EEQQEVPDPDT 10

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.382; Variance 28.285; scale 0.756

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	65	98.5	754	1	BABOH	1.74e-03
2	65	98.5	757	2	peptide-aspartate bet	1.74e-03
3	55	83.3	816	2	aspartyl beta-hydroxy	2.87e-01
4	49	74.2	226	2	probable membrane pro	4.95e+00
5	47	71.2	147	2	prolactin - golden ha	1.22e+01
6	46	69.7	68	2	single stranded DNA-b	1.94e+01
7	46	69.7	167	2	heat shock protein X4	1.91e+01
8	46	69.7	360	2	hypothetical protein	1.91e+01
9	46	69.7	704	2	probable G-box bindin	1.91e+01
10	46	69.7	7962	2	hypothetical protein	1.91e+01
11	45	68.2	217	2	elastic titin - human	2.96e+01
12	45	68.2	217	2	neuronal UNC-119 prot	2.96e+01
13	45	68.2	852	2	flagellar antigen - T	2.96e+01
14	45	68.2	3144	2	histidine-rich calciu	2.96e+01
15	45	68.2	3924	2	Huntington disease-as	2.96e+01
16	44	66.7	224	2	ankyrin 2, neuronal 1	4.55e+01
17	44	66.7	354	2	hypothetical protein	4.55e+01
18	44	66.7	411	2	uroporphyrinogen deca	4.55e+01
19	44	66.7	414	2	hypothetical protein	4.55e+01
20	44	66.7	700	1	nucleolar protein NOP	4.55e+01
21	43	65.2	140	2	transforming protein	6.95e+01
22	43	65.2	324	2	alpha-atrial natriure	6.95e+01
23	43	65.2	370	2	coat protein - phase	6.95e+01
					arrestin - bovine	6.95e+01

24 43 65.2 404 2 A28404 S-antigen - bovine 6.95e+01  
25 43 65.2 405 2 A30357 retinal S-antigen - h 6.95e+01  
26 43 65.2 412 2 A39489 transforming growth f 6.95e+01  
27 43 65.2 414 2 A31249 transforming growth f 6.95e+01  
28 43 65.2 414 1 WFKB2 transforming growth f 6.95e+01  
29 43 65.2 525 2 JN0059 hypothetical 57.4K pr 6.95e+01  
30 43 65.2 529 2 S57873 pendulin - mouse 6.95e+01  
31 43 65.2 529 2 S57345 m-importin (nuclear p 6.95e+01  
32 43 65.2 529 2 A56516 nuclear localization 6.95e+01  
33 43 65.2 572 1 HNN274 hemagglutinin-neurami 6.95e+01  
34 43 65.2 572 1 HNN273 hemagglutinin-neurami 6.95e+01  
35 43 65.2 733 2 B40595 methylmalonyl-CoA mut 6.95e+01  
36 43 65.2 767 2 T00360 hypothetical protein 6.95e+01  
37 43 65.2 851 2 T00374 hypothetical protein 6.95e+01  
38 43 65.2 855 2 A48168 proliferating-cell nu 6.95e+01  
39 43 65.2 1012 2 I53172 RAE-28 - mouse 6.95e+01  
40 43 65.2 1036 2 J53601 protein P200 - Mycopl 6.95e+01  
41 43 65.2 1281 2 J53668 dynactin 1 - mouse 6.95e+01  
42 43 65.2 1325 2 S16129 dynein-associated pro 6.95e+01  
43 43 65.2 1735 2 A57607 Munc13-1 - rat 6.95e+01  
44 43 65.2 2944 2 A54849 collagen alpha 1(VII) 6.95e+01  
45 43 65.2 3473 2 S27927 polyprotein - rice tu 6.95e+01

## ALIGNMENTS

RESULT 1  
ENTRY BABOH #type complete  
TITLE peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine  
ALTERNATE\_NAMES aspartyl (asparaginyl) beta-hydroxylase  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 31-Dec-1993 #sequence\_revision 10-Feb-1995 #text\_change 29-May-1998

ACCESSIONS A42969; A39470; B39470; C39470; S27948  
REFERENCE A42969  
#authors Jia, S.; Vandusen, W.J.; Diehl, R.E.; Kohl, N.E.; Dixon, R.A.; Elliston, K.O.; Stern, A.M.; Friedman, P.A.  
#journal J. Biol. Chem. (1992) 267:14322-14327  
#title cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.  
#cross-references EMBL:M91213; NID:g162693; PID:g162694  
#molecule\_type mRNA  
#residues 1-754 #label JIA  
#experimental\_source brain  
#note sequence extracted from NCBI backbone (NCBIP:108534)

REFERENCE A39470  
#authors Wang, Q.; Vandusen, W.J.; Petroski, C.J.; Garsky, V.M.; Stern, A.M.; Friedman, P.A.  
#journal J. Biol. Chem. (1991) 266:14004-14010  
#title Bovine liver aspartyl beta-hydroxylase. Purification and characterization.  
#cross-references M91213; NID:g162693; PID:g162694  
#accession A39470  
#molecule\_type protein  
#residues 289-328 #label WAN  
#accession B39470  
#molecule\_type protein  
#residues 615-'X', 617-630, 'XX', 633-634, 'X', 636, 'XX', 639-641  
#accession C39470  
#molecule\_type protein  
#residues 311-347, 'X', 349, 'X', 351-373, 'X', 375-379, 'X', 381-382  
#label WA3

COMMENT This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the peptidyl-aspartate substrate converts alpha-ketoglutarate to succinate and releases carbon dioxide.  
COMMENT Aspartic acid and asparagine residues in the EGF homology domain of certain plasma proteins serve as the peptidyl-aspartate substrate.  
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase; tetraatricopeptide repeat homology

```
KEYWORDS      glycoprotein; oxidoreductase; transmembrane protein
FEATURE
2-56          #domain intracellular #status predicted #label INC\
57-78         #domain transmembrane #status predicted #label TRM\
289-754       #product peptide-aspartate beta-dioxygenase, 56K form
311-754       #status predicted #label 56K\
337-370       #domain tetratricopeptide repeat homology #label TTL\
371-404       #domain tetratricopeptide repeat homology #label TT2\
13,96,466,702 #binding_site carbohydrate (Asn) (covalent) #status
              predicted
SUMMARY      #length 754 #molecular-weight 84998 #checksum 9667

Query Match      98.5%; Score 65; DB 1; Length 754;
Best Local Similarity 90.0%; Pred. NO. 1.74e-03;
Matches          9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 300 EEOQEVPPET 309
QY 1 EEOQEVPPDT 10
|||||
RESULT 2
ENTRY 138423 #type complete
TITLE aspartyl beta-hydroxylase - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change
10-Jul-1998
ACCESSIONS 138423
REFERENCE 138423
#authors Koriath, F.; Gieffers, C.; Frey, J.
#journal Gene (1994) 150:395-399
#title Cloning and characterization of the human gene encoding
         aspartyl beta-hydroxylase.
#cross-references MUID:95121937
#accession 138423
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-757 #label RES
#cross-references EMBL:U03109; NID:g458031; PID:g458032
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase;
               tetratricopeptide repeat homology
FEATURE
54-75         #domain transmembrane #status predicted #label TRM
SUMMARY      #length 757 #molecular-weight 85498 #checksum 2143

Query Match      98.5%; Score 65; DB 2; Length 757;
Best Local Similarity 90.0%; Pred. NO. 1.74e-03;
Matches          9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 304 EEOQEVPPET 313
QY 1 EEOQEVPPDT 10
|||||
RESULT 3
ENTRY S54518 #type complete
TITLE probable membrane protein YMR160w - yeast (Saccharomyces
        cerevisiae)
ALTERNATE_NAMES hypothetical protein YMR520.09
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
12-Dec-1997
ACCESSIONS S54518; S54605
REFERENCE S54510
#authors Hunt, S.; Bowman, S.
#submission submitted to the EMBL Data Library, May 1995
#accession S54518
#molecule_type DNA
#residues 1-816 #label HUN
#cross-references GB:249705; EMBL:249700; NID:g825556; PID:g825565;
        EMBL:249705; MIPS:YMR160w

GENETICS      #experimental_source strain AB972
KEYWORDS      #map_position 13R
               transmembrane protein
FEATURE
328-344       #domain transmembrane #status predicted #label TMM
SUMMARY      #length 816 #molecular-weight 95096 #checksum 3162

Query Match      83.3%; Score 55; DB 2; Length 816;
Best Local Similarity 77.8%; Pred. NO. 2.87e-01;
Matches          7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 24 EEOQEVPPQ 32
QY 1 EEOQEVPPD 9
|||||
RESULT 4
ENTRY A49159 #type complete
TITLE prolactin - golden hamster
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
16-Feb-1997
ACCESSIONS A49159
REFERENCE A49159
#authors Southard, J.N.; Sanchez-Jimenez, F.; Campbell, G.T.;
        Talamantes, F.
#journal Endocrinology (1991) 129:2965-2971
#title Sequence and expression of hamster prolactin and growth
        hormone messenger RNAs.
#cross-references MUID:92063850
#accession A49159
#status preliminary
#molecule_type mRNA
#residues 1-226 #label SOU
#note sequence extracted from NCBI backbone (NCBIN:66296,
        NCBI:P:66298)
CLASSIFICATION #superfamily prolactin
FEATURE
33-38,85-201,
218-226       #disulfide_bonds #status predicted
SUMMARY      #length 226 #molecular-weight 25582 #checksum 3394

Query Match      74.2%; Score 49; DB 2; Length 226;
Best Local Similarity 66.7%; Pred. NO. 4.95e+00;
Matches          6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 97 EEOQVPPE 105
QY 1 EEOQEVPPD 9
|||||
RESULT 5
ENTRY H70305 #type complete
TITLE single stranded DNA-binding protein - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
16-Dec-1998
ACCESSIONS H70305
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
        Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;
        Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
        J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
        Aquifex aeolicus.
#cross-references MUID:98196666
#accession H70305
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
#residues 1-147 #label AOF
```

```
##cross-references GB:AE000672; NID:g2982810; PID:g2982816; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene ssb
CLASSIFICATION #superfamily bacterial single-stranded DNA-binding protein;
single-stranded DNA-binding protein homology
FEATURE
18-96 #domain single-stranded DNA-binding protein homology
#label SSD
SUMMARY #length 147 #molecular-weight 17132 #checksum 2119
Query Match 71.2%; Score 47; DB 2; Length 147;
Best Local Similarity 75.0%; Pred. No. 1.22e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 115 EEEEEVPP 122
QY 1 EEQEVPP 8
||:|||||
RESULT 6
ENTRY C22175 #type fragment
TITLE heat shock protein X4 - African clawed frog (fragment)
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change
24-Sep-1998
ACCESSIONS C22175
REFERENCE A22175
#authors Bienz, M.
#journal Proc. Natl. Acad. Sci. U.S.A. (1984) 81:3138-3142
#title Developmental control of the heat shock response in Xenopus.
#cross-references MUID:84221917
#accession C22175
##status preliminary
##molecule_type mRNA
##residues 1-68 #label BIE
##cross-references GB:K02305; NID:g214266; PID:g214267
CLASSIFICATION #superfamily alpha-crystallin
SUMMARY #length 68 #checksum 9585
Query Match 69.7%; Score 46; DB 2; Length 68;
Best Local Similarity 55.6%; Pred. No. 1.91e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 48 DAQEIPPD 56
QY 2 EEQEVPPDT 10
||:|||||
RESULT 7
ENTRY S50808 #type complete
TITLE hypothetical protein YJL065c - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES J1115
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 13-Jan-1995 #sequence_revision 08-Sep-1995 #text_change
17-Mar-1999
ACCESSIONS S50808; S47127; S56839
REFERENCE S50798
#authors Vandenbol, M.; Durand, P.; Dion, C.; Portetelle, D.; Hilger, F.
#journal Yeast (1995) 11:57-60
#title Sequence of a 17.1 kb DNA fragment from chromosome X of Saccharomyces cerevisiae includes the mitochondrial ribosomal protein L8.
#cross-references MUID:95282514
#accession S50808
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-167 #label VAN
##cross-references EMBL:234288; NID:g498992; PID:g499003
##note the nucleotide sequence was submitted to the EMBL Data Library, June 1994
```

```
REFERENCE S47117
#authors Vandenbol, M.; Durand, P.; Dion, C.; Portetelle, D.; Hilger, F.
#submission submitted to the EMBL Data Library, June 1994
#description Sequence analysis of a 17.1 kb DNA fragment from chromosome X of Saccharomyces cerevisiae includes the mitochondrial ribosomal protein L8.
#accession S47127
##molecule_type DNA
##residues 1-167 #label VAW
##cross-references EMBL:234288; NID:g498992; PID:g499003
REFERENCE S56835
#authors Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
#submission submitted to the Protein Sequence Database, September 1995
#accession S56839
##molecule_type DNA
##residues 1-167 #label POH
##cross-references EMBL:249340; NID:g1008212; PID:g1008213; MIPS:YJL065c
GENETICS
#map_position 10L
SUMMARY #length 167 #molecular-weight 18792 #checksum 9768
Query Match 69.7%; Score 46; DB 2; Length 167;
Best Local Similarity 62.5%; Pred. No. 1.91e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 68 QQQQQVPP 75
QY 1 EEQEVPP 8
||:|||||
RESULT 8
ENTRY T03373 #type complete
TITLE probable G-box binding factor 8 - rice
ORGANISM #formal_name Oryza sativa #common_name rice
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
24-Mar-1999
ACCESSIONS T03373
REFERENCE Z14906
#authors Nakagawa, H.; Ohmiya, K.; Hattori, T.
#journal Plant J. (1996) 9:217-227
#title A rice bZIP protein, designated OSB28, is rapidly induced by abscisic acid.
#accession T03373
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-360 #label NAK
##cross-references EMBL:U42208; NID:g1147631; PID:g1147632
##experimental_source cv. Nipponbare
SUMMARY #length 360 #molecular-weight 38583 #checksum 967
Query Match 69.7%; Score 46; DB 2; Length 360;
Best Local Similarity 75.0%; Pred. No. 1.91e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 298 DEAQEVPP 305
QY 1 EEQEVPP 8
||:|||||
RESULT 9
ENTRY T02558 #type complete
TITLE hypothetical protein T26B15.15 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress
DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
05-Mar-1999
ACCESSIONS T02558
REFERENCE Z14284
#authors Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandom, R.C.; Sykes, S.M.; Mason, T.M.; Karlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
#submission submitted to the EMBL Data Library, July 1998
```

```

#description Arabidopsis thaliana chromosome II BAC T26B15 genomic
sequence.
#accession T02558
#status preliminary; translated from GB/EMBL/DDBJ
##molecule_type DNA
##residues 1-704 ##label ROU
##cross-references EMBL:AC004681; NID:g3298530; PID:g3298547
GENETICS
#map_position II
#introns 89/3; 123/3; 163/1; 184/3; 205/1; 400/3; 449/1; 543/1; 594/3;
631/3
#note T26B15.15
#length 704 #molecular-weight 79017 #checksum 2387
SUMMARY
Query Match 69.7%; Score 46; DB 2; Length 704;
Best Local Similarity 50.0%; Pred. No. 1.91e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 638 EHQEQVPPDT 10
QY 1 EHQEQVPPDT 10
RESULT 10
ENTRY 138346 #type fragment
TITLE elastic titin- human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change
05-Jun-1998
ACCESSIONS 138346
REFERENCE A57430
#authors Labat, S.; Kolmerer, B.
#journal Science (1995) 270:293-296
#title Titins: giant proteins in charge of muscle ultrastructure and
elasticity.
#cross-references MUID:96026330
#accession 138346
#status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-7962 ##label RES
##cross-references EMBL:X90569; NID:g1017426; PID:g1017427
GENETICS
#gene GDB:TTN
#cross-references GDB:127867; OMIM:188840
#map_position 2q31-2q31
#length 7962 #checksum 120
SUMMARY
Query Match 69.7%; Score 46; DB 2; Length 7962;
Best Local Similarity 62.5%; Pred. No. 1.91e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 6815 EEEVPE 6822
QY 2 EHQEQVPPD 9
RESULT 11
ENTRY JC5728 #type complete
TITLE neuronal UNC-119 protein - Caenorhabditis briggsae
ORGANISM #formal_name Caenorhabditis briggsae
DATE 09-Dec-1997 #sequence_revision 09-Dec-1997 #text_change
26-Feb-1998
ACCESSIONS JC5728
REFERENCE JC5728
#authors Maduro, M.; Pilgrim, D.
#journal Gene (1996) 183:77-85
#title Conservation of function and expression of unc-119 from two
Caenorhabditis species despite divergence of non-coding
DNA.
#cross-references MUID:97149282
#accession JC5728
##molecule_type DNA
##residues 1-217 ##label MAD

```

```

##cross-references GB:U45326; NID:gl181702; PID:gl181703
COMMENT This protein is involved in nervous system function.
GENETICS
#gene unc-119
#introns 13/3; 56/2; 173/1
SUMMARY #length 217 #molecular-weight 25094 #checksum 7052
Query Match 68.2%; Score 45; DB 2; Length 217;
Best Local Similarity 50.0%; Pred. No. 2.96e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 4 EQQQSIIPGS 13
QY 1 EQQQVPPDT 10
RESULT 12
ENTRY S47436 #type fragment
TITLE flagellar antigen - Trypanosoma brucei (fragment)
ORGANISM #formal_name Trypanosoma brucei
DATE 13-Jan-1995 #sequence_revision 30-Jan-1998 #text_change
30-Jan-1998
ACCESSIONS S47436
REFERENCE S47436
#authors Imboden, M.; Mueller, N.; Hemphill, A.; Mattioli, X.Y.2.;
Seebeck, T.
#submission submitted to the EMBL Data Library, August 1994
#description Repetitive proteins from the flagellar cytoskeleton of
African Trypanosomes are diagnostically useful antigens.
#accession S47436
##molecule_type mRNA
##residues 1-411 ##label IMB
##cross-references EMBL:Z36281; NID:g530358; PID:g530359
##experimental_source strain stock TREU 1285
SUMMARY #length 411 #checksum 428
Query Match 68.2%; Score 45; DB 2; Length 411;
Best Local Similarity 50.0%; Pred. No. 2.96e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 10 EEPQOVPAEA 19
QY 1 EQQQVPPDT 10
RESULT 13
ENTRY A34373 #type complete
TITLE histidine-rich calcium-binding protein precursor - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change
10-Sep-1997
ACCESSIONS A34373
REFERENCE A34373
#authors Hofmann, S.L.; Goldstein, J.L.; Orth, K.; Moomaw, C.R.;
Slaughter, C.A.; Brown, M.S.
#journal J. Biol. Chem. (1989) 264:18083-18090
#title Molecular cloning of a histidine-rich Ca(2+)-binding protein
of sarcoplasmic reticulum that contains highly conserved
repeated elements.
#cross-references MUID:90036884
#accession A34373
#status preliminary
##molecule_type mRNA
##residues 1-852 ##label HOF
##cross-references GB:J05080; NID:gl65099; PID:gl65100
KEYWORDS calcium binding
SUMMARY #length 852 #molecular-weight 96116 #checksum 3434
Query Match 68.2%; Score 45; DB 2; Length 852;
Best Local Similarity 55.6%; Pred. No. 2.96e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 190 EEEEEVSPE 198
||:||||:
QY 1 EQQEVPPD 9

RESULT 14
ENTRY A46068 #type complete
TITLE Huntingdon disease-associated protein - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
29-Aug-1997
A46068; 154337
A46068;
#authors MacDonald, M.E.; Ambrose, C.M.; Duvao, M.P.; Myers, R.H.;
Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S.A.; James, M.;
Groot, N.; MacFarlane, H.; Jenkins, B.; Anderson, M.A.;
Wexler, N.S.; Gusella, J.F.; Bates, G.P.; Baxendale, S.;
Hummerich, H.; Kirby, S.; North, M.; Youngman, S.; Mott,
R.; Zehetner, G.; Sedlacek, Z.; Poustka, A.; Frischauf,
A.M.; Buckler, A.J.; Church, D.; Doucette-Stamm, L.;
O'Donovan, M.C.; Riba-Ramirez, L.; Shah, M.; Stanton, V.P.;
Strobel, S.A.; Draths, K.M.; Wales, J.L.; Dervan, P.;
Housman, D.E.; Altherr, M.; Shiang, R.; Thompson, L.;
Fielder, T.; Wasnuth, J.J.; Tagle, D.; Valdes, J.; Elmer,
L.; Allard, M.; Castilla, L.; Swaroop, M.; Blanchard, K.;
Collins, F.S.; Snell, R.; Holloway, T.; Gillespie, K.;
Datson, N.; Shaw, D.; Harper, P.S.
#journal Cell (1993) 72:971-983
#title A novel gene containing a trinucleotide repeat that is
expanded and unstable on Huntington's disease chromosomes.
#accession A46068
#status preliminary
#molecule_type mRNA
#residues 1-3144 #label MAC
#cross-references GB:L12392
REFERENCE 154337
#authors Lin, B.; Rommens, J.M.; Graham, R.K.; Kalchman, M.;
MacDonald, H.; Nasir, J.; Delaney, A.; Goldberg, Y.P.;
Hayden, M.R.
Hum. Mol. Genet. (1993) 2:1541-1545
#journal Differential 3' polyadenylation of the Huntington disease
#title gene results in two mRNA species with variable tissue
expression.
#cross-references MUID:94093536
#accession 154337
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 2563-3144 #label RES
#cross-references GB:L20431; NID:g398028; PID:g398029
GENETICS
#gene GDB:HD
#cross-references GDB:l19307; OMIM:143100
#map_position 4p16.3-4p16.3
SUMMARY #length 3144 #molecular_weight 347896 #checksum 1801

Query Match 58.2%; Score 45; DB 2; Length 3144;
Best Local Similarity 60.0%; Pred. No. 2.96e-01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 EEEEEVDNPT 2352
||:||||:
QY 1 EQQEVPPDT 10

RESULT 15
ENTRY S37431 #type complete
TITLE ankyrin 2, neuronal long splice form - human
ALTERNATE_NAMES ankyrin B, 440K splice form; ankyrin-B; brain ankyrin;
non-erythroid ankyrin
CONTAINS ankyrin 2, short form
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
10-Jul-1998

ACCESSIONS S37431; A39643; B39643; A40334; A49462; S14533; S14569
REFERENCE S37431
#authors Chan, W.
#submission submitted to the EMBL Data Library, September 1993
#accession S37431
#status preliminary
#molecule_type mRNA
#residues 1-3924 #label CHA
#cross-references EMBL:Z26634; NID:g406287; PID:g406288
REFERENCE A39643
#authors Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.
#journal J. Cell Biol. (1991) 114:241-253
#title Isolation and characterization of cDNAs encoding human brain
ankyrins reveal a family of alternatively spliced genes.
#cross-references MUID:91302466
#accession A39643
#status preliminary
#molecule_type mRNA
#residues 1-2077 #label OT1
#cross-references GB:X56957
#accession B39643
#status preliminary
#molecule_type mRNA
#residues 1-1443,3585-3924 #label OTT
#cross-references EMBL:X56958
REFERENCE A40334
#authors Tse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.;
Sahr, K.E.; Lux, S.E.; Ward, D.C.; Forget, B.G.
#journal Genomics (1991) 10:858-866
#title Isolation and chromosomal localization of a novel
nonerythroid ankyrin gene.
#cross-references MUID:92009921
#accession A40334
#status preliminary
#molecule_type DNA
#residues 463-474, 'PE', 477-495 #label TSE
#cross-references GB:M37123; NID:g178647; PID:g178648
REFERENCE A49462
#authors Chan, W.; Kordeli, E.; Bennett, V.
#journal J. Cell Biol. (1993) 123:1463-1473
#title 440-Kb ankyrinB: structure of the major developmentally
regulated domain and selective localization in unmyelinated
axons.
#cross-references MUID:94075409
#accession A49462
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-3924 #label RES
#cross-references EMBL:Z26634; NID:g406287; PID:g406288
GENETICS
#gene GDB:ANK2
#cross-references GDB:127607; OMIM:106410
#map_position 4q25-4q27
CLASSIFICATION #superfamily ankyrin; ankyrin repeat homology
KEYWORDS alternative splicing
FEATURE
2-3924 #product ankyrin 2, long form #status predicted #label
MAT\
#product ankyrin 2, short form #status predicted #label
MA2\
63-95 #domain ankyrin repeat homology #label AN01\
96-128 #domain ankyrin repeat homology #label AN02\
129-161 #domain ankyrin repeat homology #label AN03\
162-190 #domain ankyrin repeat homology #label AN04\
191-223 #domain ankyrin repeat homology #label AN05\
232-264 #domain ankyrin repeat homology #label AN06\
265-297 #domain ankyrin repeat homology #label AN07\
298-330 #domain ankyrin repeat homology #label AN08\
331-363 #domain ankyrin repeat homology #label AN09\
364-396 #domain ankyrin repeat homology #label AN10\
397-429 #domain ankyrin repeat homology #label AN11\
430-462 #domain ankyrin repeat homology #label AN12\
463-495 #domain ankyrin repeat homology #label AN13\
496-528 #domain ankyrin repeat homology #label AN14\

```

529-561 #domain ankylrin repeat homology #label AN15\  
 562-594 #domain ankylrin repeat homology #label AN16\  
 595-627 #domain ankylrin repeat homology #label AN17\  
 628-660 #domain ankylrin repeat homology #label AN18\  
 661-693 #domain ankylrin repeat homology #label AN19\  
 694-726 #domain ankylrin repeat homology #label AN20\  
 727-759 #domain ankylrin repeat homology #label AN21\  
 760-792 #domain ankylrin repeat homology #label AN22\  
 793-825 #domain ankylrin repeat homology #label AN23\  
 SUMMARY #length 3924 #molecular-weight 430340 #checksum 3664

Query Match 68.2%; Score 45; DB 2; Length 3924;  
 Best Local Similarity 40.0%; Pred. No. 2.96e+01;  
 Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 3839 DDMPEIPPET 3848

QY 1 EEQQEVPPT 10

Search completed: Thu Oct 21 15:37:08 1999  
 Job time : 13 secs.

\*\*\*\*\*  
[W][O][R][L][D]  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 21 15:37:24 1999; MasPar time 2.30 Seconds  
Tabular output not generated. 122.893 Million cell updates/sec

Title: >US-09-040-485-7  
Description: (1-10) from US09040485.pep  
Perfect Score: 66  
Sequence: 1 EQQQEVPPDT 10

Scoring table: PAM 150  
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 22.120; Variance 25.762; scale 0.859

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	65	98.5	734	1 ASPH_BOVIN ASPARTYL/ASPARAGINYL B 2.81e-04
2	65	98.5	757	1 ASPH_HUMAN ASPARTYL/ASPARAGINYL B 2.81e-04
3	55	83.3	816	1 YM35_YEAST HYPOTHETICAL 95.1 KD P 8.10e-02
4	49	74.2	226	1 PRL_MESAU PROLACTIN PRECURSOR (P 1.88e+00
5	48	72.7	815	1 LU15_HUMAN PUTATIVE TUMOR SUPPRES 3.10e+00
6	46	69.7	167	1 YJG5_YEAST HYPOTHETICAL 18.8 KD P 8.27e+00
7	46	69.7	775	1 LYS4_EMENI HOMOACONITASE PRECURSO 8.27e+00
8	45	68.2	217	1 U119_CABER UNC-119 PROTEIN. 1.34e+01
9	45	68.2	671	1 PEXH_YARLI PEROXISOMAL MEMBRANE P 1.34e+01
10	45	68.2	852	1 SRCH_RABIT SARCOPLASMIC RETICULUM 1.34e+01
11	45	68.2	1839	1 ANKC_HUMAN ANKYRIN, BRAIN VARIANT 1.34e+01
12	45	68.2	3144	1 HD_HUMAN HUNTINGTIN (HUNTINGTON 1.34e+01
13	45	68.2	3924	1 ANKB_HUMAN ANKYRIN, BRAIN VARIANT 1.34e+01
14	44	66.7	224	1 YKEL_YEAST HYPOTHETICAL 26.2 KD P 2.14e+01
15	44	66.7	354	1 DCUP_ECOLI UROPOHRYRINOGEN DECAR 2.14e+01
16	44	66.7	414	1 NOP3_YEAST NUCLEOLAR PROTEIN 3 (M 2.14e+01
17	44	66.7	700	1 MYB_HUMAN MYB-RELATED PROTEIN B 2.14e+01
18	44	66.7	755	1 RREL_HUMAN RAS-RESPONSIVE ELEMENT 2.14e+01
19	43	65.2	140	1 ANF_CHICK ATRIAL NATRIURETIC FAC 3.40e+01
20	43	65.2	184	1 K501_ACTCH FRUIT PROTEIN PKIWI501 3.40e+01
21	43	65.2	395	1 G52A_CHICK G2/MITOTIC-SPECIFIC CY 3.40e+01
22	43	65.2	404	1 ARRS_BOVIN S-ARRESTIN (RETINAL S- 3.40e+01
23	43	65.2	405	1 ARRS_HUMAN S-ARRESTIN (RETINAL S- 3.40e+01



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M91213; G162694; -  
 KW OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;  
 KW ENDOPLASMIC RETICULUM  
 FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 58 78 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT DOMAIN 79 754 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9 12 POLY-GLY.  
 FT DOMAIN 14 21 POLY-SER.  
 FT DOMAIN 318 328 POLY-LYS.  
 FT CARBOHYD 96 96 POTENTIAL.  
 FT CARBOHYD 466 466 POTENTIAL.  
 FT CARBOHYD 702 702 POTENTIAL.  
 SQ SEQUENCE 754 AA; 84998 MW; 608861B2 CRC32;

Query Match 98.5%; Score 65; DB 1; Length 754;  
 Best Local Similarity 90.0%; Pred. No. 2.81e-04;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 300 EEOQEVPPET 309  
 QY 1 EEOQEVPPDT 10  
 |||||

## RESULT 2

ID ASPH HUMAN STANDARD; PRT; 757 AA.  
 AC Q12797;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-DIOXYGENASE).  
 DE ASPH.  
 GN HOMO SAPIENS (HUMAN).  
 OS EURARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95121937.  
 RA KORIOTH F., GIEFFERS C., FREY J.;  
 RT "Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase."  
 RL GENE 150:395-399(1994).  
 CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.  
 CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) = PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).  
 CC -!- COFACTOR: IRON.  
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.  
 CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.  
 CC -!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U03109; G458032; -  
 DR MIM; 600582; -  
 KW OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;

KW ENDOPLASMIC RETICULUM.  
 FT DOMAIN 1 54 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 55 75 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT DOMAIN 76 757 LUMENAL (POTENTIAL).  
 FT TRANSMEM 13 20 POLY-SER.  
 FT DOMAIN 323 332 POLY-LYS.  
 FT CARBOHYD 452 452 POTENTIAL.  
 FT CARBOHYD 705 705 POTENTIAL.  
 SQ SEQUENCE 757 AA; 85498 MW; A6GAF24 CRC32;

Query Match 98.5%; Score 65; DB 1; Length 757;  
 Best Local Similarity 90.0%; Pred. No. 2.81e-04;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 304 EEOQEVPPET 313  
 QY 1 EEOQEVPPDT 10  
 |||||

## RESULT 3

ID YM35 YEAST STANDARD; PRT; 816 AA.  
 AC Q03823;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 95.1 KD PROTEIN IN IMP1-HLJ1 INTERGENIC REGION.  
 GN YMR160W OR YMR520.09.  
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EURARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
 OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA HUNT S., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.;  
 RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; 249705; G825565; -  
 KW HYPOTHETICAL PROTEIN  
 SQ SEQUENCE 816 AA; 95096 MW; A897E595 CRC32;

Query Match 83.3%; Score 55; DB 1; Length 816;  
 Best Local Similarity 77.8%; Pred. No. 8.10e-02;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 24 EEOQEVPPQ 32  
 QY 1 EEOQEVPPD 9  
 |||||

## RESULT 4

ID PRL MESAU STANDARD; PRT; 226 AA.  
 AC P37884;  
 DT 01-OCT-1994 (REL. 30, CREATED)  
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE PROLACTIN PRECURSOR (PRL).  
 GN PRL.  
 OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).  
 OC EURARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; CRICETINAE; MESOCRICETUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92063850.  
 RA SOUTHARD J.N., SANCHEZ-JIMENEZ F., CAMPBELL G.T., TALAMANTES F.;

RT "sequence and expression of hamster prolactin and growth hormone  
messenger RNAs.";  
RL ENDOCRINOLOGY 129:2965-2971(1991).  
CC -!- FUNCTION: PROLACTIN ACTS PRIMARILY ON THE MAMMARY GLAND BY  
CC PROMOTING LACTATION.  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; S66296; G239354; .  
DR PIR; A49159; A49159.  
DR PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
DR PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
DR PFAM; PF00103; hormone; 1.  
DR HSP; Q28632; IAN3.  
DR HORMONE; PARTURITION; LACTATION; PITUITARY; SIGNAL.  
KW SIGNAL 1 29 BY SIMILARITY.  
FT CHAIN 1 29 PROLACTIN.  
FT DISULFID 33 38 BY SIMILARITY.  
FT DISULFID 85 201 BY SIMILARITY.  
FT DISULFID 218 226 BY SIMILARITY.  
SQ SEQUENCE 226 AA; 25582 MW; D5A4C5BC CRC32;  
  
Query Match 74.2%; Score 49; DB 1; Length 226;  
Best Local Similarity 66.7%; Pred. No. 1.88e+00;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Db 97 EEAQVPPPE 105  
QY 1 EEQVEVPPD 9  
  
RESULT 5  
ID LU15\_HUMAN STANDARD; PRT; 815 AA.  
AC P52756;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE PUTATIVE TUMOR SUPPRESSOR LUCA15.  
GN LUCA15.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BADER S., LATIF F., DUH F., WEI M., KASHUBA V., SEKIDO Y., LEE C.,  
RA KOONIN E., ZABAROFKY E., KLEIN G., MINNA J.D., LERNAN M.;  
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RNP).  
CC -!- SIMILARITY: HIGH, TO HUMAN DXS237E.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U23946; G1244404; .  
DR PROSITE; PS00030; RNP\_1; 2.  
DR PFAM; PF00076; rim; 2.  
DR PFAM; PF00641; zf-RanBP; 1.  
KW ANTI-ONCOGENE; RNA-BINDING; NUCLEAR PROTEIN.  
FT DOMAIN 140 147 RNA-BINDING (RNPI) (BY SIMILARITY).

FT DOMAIN 274 281 RNA-BINDING (RNPI) (BY SIMILARITY).  
SQ SEQUENCE 815 AA; 92073 MW; 30EF5EB2 CRC32;  
  
Query Match 72.7%; Score 48; DB 1; Length 815;  
Best Local Similarity 60.0%; Pred. No. 3.10e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Db 210 DSEQVEPPGT 219  
QY 1 EEQVEVPPDT 10  
  
RESULT 6  
ID YJG5\_YEAST STANDARD; PRT; 167 AA.  
AC P40366;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 18.8 KD PROTEIN IN SMC3-MRPL8 INTERGENIC REGION.  
GN YJL065C OR J1115 OR HRD167.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE; 95282514.  
RA VANDENBOL M., DURAND P., DION C., PORTELELE D., HILGER F.;  
RT "Sequence of a 17.1 kb DNA fragment from chromosome X of  
RT Saccharomyces cerevisiae includes the mitochondrial ribosomal protein  
RT L8.";  
RL YEAST 11:57-60(1995).  
CC -!- SIMILARITY: TO YEAST DNA POLYMERASE EPSILON, SUBUNIT C (DPB3).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; 234288; G499003; .  
DR EMBL; 249340; G1008213; .  
DR PIR; S47127; S47127.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 167 AA; 18792 MW; B5DC5E3C CRC32;  
  
Query Match 69.7%; Score 45; DB 1; Length 167;  
Best Local Similarity 62.5%; Pred. No. 8.27e+00;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
Db 68 QOQOQVPP 75  
QY 1 EEQVEVPP 8  
  
RESULT 7  
ID LYS4\_EMENT STANDARD; PRT; 775 AA.  
AC Q92412;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE HOMOACONITASE PRECURSOR (EC 4.2.1.36) (HOMOACONITATE HYDRATASE).  
GN LYSF.  
OS EMERICELLA NIDULANS (ASPERGILLUS NIDULANS).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;  
OC EURHOTIALES; TRICHOCOMACEAE; EMERICELLA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97411901.  
RA WEIDNER G., STEFFAN B., BRAKHAGE A.A.;  
RT "The Aspergillus nidulans lysf gene encodes homoacnitate, an enzyme

involved in the fungus-specific lysine biosynthesis pathway.";  
RL MOL. GENET. 255:237-247(1997).  
CC -1- FUNCTION: RESPONSIBLE FOR THE DEHYDRATION OF CIS-HOMOACONITATE TO  
CC HOMOSOCITRIC ACID.  
CC -1- CATALYTIC ACTIVITY: 2-HYDROXYBUTANE-1,2,4-TRICARBOXYLATE = BUT-1-  
CC ENH-1,2,4-TRICARBOXYLATE + H(2)O.  
CC -1- PATHWAY: THIRD STEP IN LYSINE BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X96624; E265392; -  
DR PROSITE; PS00450; ACONITASE\_1; 1.  
DR PROSITE; PS01244; ACONITASE\_2; 1.  
DR PFAM; PF00330; aconitase; 2.  
DR PFAM; PF00694; Aconitase\_C; 1.  
KW LYSINE BIOSYNTHESIS; LYASE; MITOCHONDRION; TRANSIT PEPTIDE;  
FT IRON-SULFUR.  
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).  
FT CHAIN ? 775 HOMOACONITASE  
FT METAL 394 394 IRON (IRON-SULFUR CLUSTER)  
FT METAL 462 462 (BY SIMILARITY).  
FT METAL 462 462 IRON (IRON-SULFUR CLUSTER)  
FT METAL 465 465 (BY SIMILARITY).  
FT METAL 465 465 IRON (IRON-SULFUR CLUSTER)  
FT METAL 465 465 (BY SIMILARITY).  
SQ SEQUENCE 775 AA; 84038 MW; 86DE640 CRC32;  
  
Query Match 69.7%; Score 46; DB 1; Length 775;  
Best Local Similarity 62.5%; Pred. No. 8.27e+00;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
Db 619 QDDVPPET 626  
Y 1:|||||  
Y 3 QOEVPPT 10  
  
RESULT 8  
ID U119\_CAEBR STANDARD; PRT; 217 AA.  
AC Q17297;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE UNC-119 PROTEIN.  
GN UNC-119.  
OS CAENORHABDITIS BRIGGSAE.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97149282.  
RA MADURO M.F., PILGRIM D.B.;  
RT "Conservation of function and expression of unc-119 from two  
RT Caenorhabditis species despite divergence of non-coding DNA.";  
RL GENE 183:77-85(1996).  
CC -1- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OR FUNCTION OF THE  
CC NERVOUS SYSTEM (BY SIMILARITY).  
CC -1- SIMILARITY: STRONG, TO MAMMALIAN RETINAL PROTEIN 4; WEAK, TO  
CC C.ELEGANS C27H5.1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; U45326; G1181703; -  
KW DEVELOPMENTAL PROTEIN.  
SQ SEQUENCE 217 AA; 25094 MW; F477F555 CRC32;  
  
Query Match 68.2%; Score 45; DB 1; Length 217;  
Best Local Similarity 50.0%; Pred. No. 1.34e+01;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
Db 4 EQQSIIPGS 13  
Y 1:|||||  
Y 1 EQQEVPPPT 10  
  
RESULT 9  
ID PEXH\_YARLI STANDARD; PRT; 671 AA.  
AC P87200;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE PEROXISOMAL MEMBRANE PROTEIN PEX17 (PEROXIN-17).  
GN PEX17.  
OS YARROWIA LIPOLYTICA (CANDIDA LIPOLYTICA).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
OC DIPODASCACEAE; YARROWIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-EI22;  
RX MEDLINE; 97265383.  
RA SMITH J.J., SZILARD R.K., MARELLI M., RACHUBINSKI R.A.;  
RT "The peroxin Pex17p of the yeast Yarrowia lipolytica is associated  
RT peripherally with the peroxisomal membrane and is required for the  
RT import of a subset of matrix proteins.";  
RL MOL. CELL. BIOL. 17:2511-2520(1997).  
CC -1- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. REQUIRED FOR THE  
CC IMPORT OF A SUBSET OF MATRIX PROTEINS.  
CC -1- SUBCELLULAR LOCATION: PEROXISOMAL MEMBRANE-ASSOCIATED.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U73028; G2039277; -  
KW PEROXISOME; TRANSMEMBRANE.  
FT TRANSMEM 127 147 POTENTIAL.  
FT TRANSMEM 187 207 POTENTIAL.  
FT TRANSMEM 212 232 POTENTIAL.  
FT TRANSMEM 258 278 POTENTIAL.  
FT TRANSMEM 318 338 POTENTIAL.  
FT TRANSMEM 372 392 POTENTIAL.  
FT TRANSMEM 475 495 POTENTIAL.  
FT TRANSMEM 504 524 POTENTIAL.  
SQ SEQUENCE 671 AA; 75588 MW; BC5618B9 CRC32;  
  
Query Match 68.2%; Score 45; DB 1; Length 671;  
Best Local Similarity 50.0%; Pred. No. 1.34e+01;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
Db 573 QOEVDPP 580  
Y 1:|||||  
Y 1 EQQEVPP 8  
  
RESULT 10  
ID SRCH\_RABIT STANDARD; PRT; 852 AA.  
AC P16230;  
DT 01-APR-1990 (REL. 14, CREATED)  
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)

DE SARCOPLASMIC RETICULUM HISTIDINE-RICH CALCIUM-BINDING PROTEIN  
DE PRECURSOR (HCP).  
OS ORYCTOLAGUS CUNICULUS (RABBIT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE-SKELETAL MUSCLE;  
RA HOFMANN S.L., GOLDSTEIN J.L., ORTH K., MOOMAW C.R., SLAUGHTER C.A.,  
RA BROWN M.S.;  
RT "Molecular cloning of a histidine-rich Ca2+-binding protein of  
sarcoplasmic reticulum that contains highly conserved repeated  
elements";  
RL J. BIOL. CHEM. 264:18083-18090(1989).  
CC -!- FUNCTION: HCP MAY PLAY A ROLE IN THE REGULATION OF CA(2+)  
CC SEQUESTRATION OR RELEASE IN THE SR OF SKELETAL AND CARDIAC  
CC MUSCLE.  
CC -!- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN.  
CC -!- SIMILARITY: STRONG TO HUMAN HRC.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J05080; G165100; -  
DR PIR; A34373; A34373.  
DR PROSITE; PS00328; HCP; 10.  
KW CALCIUM-BINDING; SIGNAL; REPEAT.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 852 HISTIDINE-RICH CALCIUM-BINDING PROTEIN.  
FT MOD\_RES 28 28 BLOCKED.  
FT DOMAIN 59 100 2 X APPROXIMATE TANDEM REPEATS.  
FT REPEAT 59 79 1-1.  
FT REPEAT 80 100 1-2.  
FT DOMAIN 199 470 10 X TANDEM REPEATS, ACIDIC.  
FT REPEAT 199 224 2-1.  
FT REPEAT 225 253 2-2.  
FT REPEAT 254 282 2-3.  
FT REPEAT 283 310 2-4.  
FT REPEAT 311 339 2-5.  
FT REPEAT 340 367 2-6.  
FT REPEAT 368 395 2-7.  
FT REPEAT 396 423 2-8.  
FT REPEAT 424 451 2-9.  
FT REPEAT 452 470 2-10.  
FT DOMAIN 471 585 4 X APPROXIMATE TANDEM REPEATS.  
FT DOMAIN 721 733 POLY-GLU.  
FT DOMAIN 780 836 METAL-BINDING (POTENTIAL).  
SQ SEQUENCE 852 AA; 96117 MW; 3C1DF781 CRC32;  
Query Match 68.28; Score 45; DB 1; Length 852;  
Best Local Similarity 55.68; Pred. No. 1.34e+01;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 190 EEEEEVSPE 198  
QY 1 EQQEVPPD 9  
RESULT 11  
ID ANKC\_HUMAN STANDARD; PRT; 1839 AA.  
AC Q01485;  
DT 01-APR-1993 (REL. 25, CREATED)  
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE ANKYRIN, BRAIN VARIANT 2 (ANKYRIN B) (ANKYRIN, NONERYTHROID)  
DE (FRAGMENT).  
GN ANK2.

OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN STEM;  
RX MEDLINE; 91302456.  
RA OTTO E., KUNIMOTO M., MCLAUGHLIN T., BENNETT V.;  
RT "Isolation and characterization of cDNAs encoding human brain  
ankyrins reveal a family of alternatively spliced genes.";  
RL J. CELL BIOL. 114:241-253(1991).  
RN [2]  
RP SEQUENCE OF 463-495 FROM N.A.  
RX MEDLINE; 92009921.  
RA TSE W.T., MENNINGER J.C., YANG-FENG T.L., FRANCKE U., SAHR K.E.,  
RA LUX S.E., WARD D.C., FORGET B.G.;  
RT "Isolation and chromosomal localization of a novel nonerythroid  
ankyrin gene.";  
RL GENOMICS 10:858-866(1991).  
CC -!- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-  
CC SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN GP85,  
CC BAND 4.2, TO NA-K ATPASE, TO THE LYMPHOCYTE MEMBRANE PROTEIN GP85,  
CC AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND  
CC DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO  
CC THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;  
CC THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.  
CC -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL  
CC CELLS THROUGHOUT THE BRAIN.  
CC -!- ALTERNATIVE PRODUCTS: THE TWO BRAIN VARIANTS ARE PRODUCED BY  
CC ALTERNATIVE SPLICING OF GENE ANK2.  
CC -!- SIMILARITY: CONTAINS 24 ANK REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X56958; G29491; -  
DR EMBL; M37123; G178648; -  
DR PIR; S14569; S14569.  
DR PIR; B39643; B39643.  
DR MIM; 106410; -  
DR PROSITE; PS50017; DEATH\_DOMAIN; 1.  
DR PFAM; PF00023; ank; 22.  
DR PFAM; PF00531; death; 1.  
DR PFAM; PF00791; ZUS; 1.  
DR HSSP; Q00420; IAWC.  
KW CYTOSKELETON; ALTERNATIVE SPLICING; REPEAT; ANK REPEAT;  
KW MULTIGENE FAMILY.  
FT DOMAIN 63 792 22 X ANK MOTIF REPEATS.  
FT REPEAT 63 95 ANK MOTIF 1.  
FT REPEAT 96 128 ANK MOTIF 2.  
FT REPEAT 129 161 ANK MOTIF 3.  
FT REPEAT 162 190 ANK MOTIF 4.  
FT REPEAT 191 231 ANK MOTIF 5.  
FT REPEAT 232 264 ANK MOTIF 6.  
FT REPEAT 265 297 ANK MOTIF 7.  
FT REPEAT 298 330 ANK MOTIF 8.  
FT REPEAT 331 363 ANK MOTIF 9.  
FT REPEAT 364 396 ANK MOTIF 10.  
FT REPEAT 397 429 ANK MOTIF 11.  
FT REPEAT 430 462 ANK MOTIF 12.  
FT REPEAT 463 495 ANK MOTIF 13.  
FT REPEAT 496 528 ANK MOTIF 14.  
FT REPEAT 529 561 ANK MOTIF 15.  
FT REPEAT 562 594 ANK MOTIF 16.  
FT REPEAT 595 627 ANK MOTIF 17.  
FT REPEAT 628 660 ANK MOTIF 18.  
FT REPEAT 661 693 ANK MOTIF 19.  
FT REPEAT 694 726 ANK MOTIF 20.

FT REPEAT 727 759 ANK MOTIF 21.  
 FT REPEAT 760 792 ANK MOTIF 22.  
 FT DOMAIN 1451 1535 DEATH DOMAIN.  
 FT CONFLICT 475 476 GQ -> PE (IN REF. 2).  
 SQ NON\_TER 1839 1839  
 SQ SEQUENCE 1839 AA; 202409 MW; 546A50B4 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 1839;  
 Best Local Similarity 40.0%; Pred. No. 1.34e+01;  
 Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 1754 DDMPEIPPT 1763  
 : : l:l:l:l  
 QY 1 EEQEVPPDT 10

## RESULT 12

ID HD\_HUMAN STANDARD; PRT; 3144 AA.  
 AC P42858;  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE HUNTINGTIN (HUNTINGTIN'S DISEASE PROTEIN) (HD PROTEIN).  
 GN HD OR IT15.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-RETINA:  
 RX MEDLINE; 93208892.  
 RA MACDONALD M., AMBROSE C.M., DUYAO M.P., MYERS R.H., LIN C.S.,  
 RA SRINIDHI J., BARNES G., TAYLOR S.A., JAMES M., GROOT N., MCFARLANE H.,  
 RA JENKINS B., ANDERSON M.A., WEXLER N.S., GUSELLA J.F., BATES G.P.,  
 RA BAXENDALE S., HUMMERICH H., KIRBY S., NORTH M., YOUNGMAN S., MOTT R.,  
 RA ZHEUTNER G., SEDLACEK Z., POUSTKA A., FRISCHAUF A.-M., LEHRACH H.,  
 RA BUCKLER A.J., CHURCH D., DOUCETTE-STAMM L., O'DONOVAN M.C.,  
 RA RIBA-RAMIREZ L., SHAH M., STANTON V.P., STROBEL S.A., DRATHS K.M.,  
 RA WALES J.L., DERVAN P., HOUSMAN D.E., ALTHERR M., SHIANG R.,  
 RA THOMPSON L., FIELDER T., WASMUTH J.J., TAGLE D., VALDES J.,  
 RA ALLARD M., CASTILLA L., SWAROOP M., BLANCHARD K., COLLINS F.S.,  
 RA SNELL R., HOLLOWAY T., GILLESPIE K., DATSON N., SHAW S., HARPER P.S.;  
 RT "A novel gene containing a trinucleotide repeat that is expanded and  
 RT unstable on Huntington's disease chromosomes. The Huntington's  
 RT Disease Collaborative Research Group.";  
 RL CELL 72:971-983(1993).  
 RN [2]  
 RP SEQUENCE OF 1-90 FROM N.A.  
 RX MEDLINE; 95278941.  
 RA LIN B., NASIR J., KALCHMAN M.A., MCDONALD H., ZEISLER J.,  
 RA GOLDBERG Y.P., HAYDEN M.R.;  
 RT "Structural analysis of the 5' region of mouse and human Huntington  
 RT disease genes reveals conservation of putative promoter region and  
 RT dl- and trinucleotide polymorphisms.";  
 RL GENOMICS 25:707-715(1995).  
 RN [3]  
 RP SEQUENCE OF 1-205 FROM N.A.  
 RX MEDLINE; 94255787.  
 RA AMBROSE C.M., DUYAO M.P., BARNES G., BATES G.P., LIN C.S.,  
 RA SRINIDHI J., BAXENDALE S., HUMMERICH H., LEHRACH H., ALTHERR M.,  
 RA WASMUTH J., BUCKLER A., CHURCH D., HOUSMAN D., BERKS M., MICKLEM G.,  
 RA DURBIN R., DODGE A., READ A., GUSELLA J.F., MACDONALD M.E.;  
 RT "Structure and expression of the Huntington's disease gene: evidence  
 RT against simple inactivation due to an expanded CAG repeat.";  
 RL SOMAT. CELL MOL. GENET. 20:27-38(1994).  
 RN [4]  
 RP SEQUENCE OF 1-117 FROM N.A.  
 RX MATTHEWS P.;  
 RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [5]  
 RP SEQUENCE OF 119-934 FROM N.A.  
 RA LLOYD C.;  
 RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [6]  
 RP SEQUENCE OF 1212-1290 FROM N.A.  
 RA MUNGALL A., ODELL C.;  
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [7]  
 RP SEQUENCE OF 1291-1860 FROM N.A.  
 RA MUNGALL A.;  
 RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [8]  
 RP SEQUENCE OF 1862-2820 FROM N.A.  
 RA BUCK D.;  
 RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [9]  
 RP SEQUENCE OF 2563-3144 FROM N.A.  
 RC TISSUE=FRONTAL CORTEX, BRAIN, RETINA, CAUDATE, AND MUSCLE;  
 RX MEDLINE; 94093536.  
 RA LIN B., ROMMENS J.M., GRAHAM R.K., KALCHMAN M., MACDONALD H.,  
 RA NASIR J., DELANEY A., GOLDBERG Y.P., HAYDEN M.R.;  
 RT "Differential 3' polyadenylation of the Huntington disease gene  
 RT results in two mRNA species with variable tissue expression.";  
 RL HUM. MOL. GENET. 2:1541-1545(1993).  
 RN [10]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE; 95375771.  
 RA TROTTIER Y., DEVYS D., IMBERT G., SAUDOU F., AN I., LUTZ Y., WEBER C.,  
 RA AGID Y., HIRSCH E.C., MANDEL J.-L.;  
 RT "Cellular localization of the Huntington's disease protein and  
 RT discrimination of the normal and mutated form.";  
 RL NAT. GENET. 10:104-110(1995).  
 RN [11]  
 RP CLEAVAGE BY APOPAIN.  
 RX MEDLINE; 96331285.  
 RA GOLDBERG Y.P., NICHOLSON D.W., RASPER D.M., KALCHMAN M.A., KOIDE H.B.,  
 RA GRAHAM R.K., BROMM M., KAZEMI-ESFARJANI P., THORBERRY N.A.,  
 RA VAILLANCOURT J.P., HAYDEN M.R.;  
 RT "Cleavage of huntingtin by apopain, a proapoptotic cysteine protease,  
 RT is modulated by the polyglutamine tract.";  
 RL NAT. GENET. 13:442-449(1996).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN MICROTUBULE-MEDIATED TRANSPORT OR  
 CC VESICLE FUNCTION.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN A VARIETY OF TISSUES WITH  
 CC THE HIGHEST LEVEL EXPRESSION IN THE BRAIN (NERVE FIBRES,  
 CC VARICOITIES, AND NERVE ENDINGS). IN THE BRAIN, THE REGIONS WHERE  
 CC IT CAN BE MAINLY FOUND ARE THE CEREBELLAR CORTEX, THE NEOCORTEX,  
 CC THE STRIATUM, AND THE HIPPOCAMPAL FORMATION.  
 CC -!- PTM: CLEAVED BY APOPAIN DOWNSTREAM OF THE POLYGLUTAMINE STRETCH.  
 CC THE RESULTING AMINO-TERMINAL FRAGMENT IS CYTOTOXIC AND PROVOKES  
 CC APOPTOSIS.  
 CC -!- POLYMORPHISM: THE POLY-GLN REGION OF HD IS HIGHLY POLYMORPHIC (10  
 CC TO 35 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO ABOUT  
 CC 36-120 REPEATS IN HD PATIENTS. THE REPEAT LENGTH USUALLY INCREASES  
 CC IN SUCCESSIVE GENERATIONS, BUT CONTRACTS ALSO ON OCCASION. THE  
 CC LONGER EXPANSIONS RESULT IN EARLIER ONSET AND MORE SEVERE CLINICAL  
 CC MANIFESTATIONS OF THE DISEASE. THE ADJACENT POLY-PRO REGION IS  
 CC ALSO POLYMORPHIC AND VARIES BETWEEN 7-12 RESIDUES. POLYGLUTAMINE  
 CC EXPANSION LEADS TO ELEVATED SUSCEPTIBILITY TO APOPAIN CLEAVAGE AND  
 CC LIKELY RESULT IN ACCELERATED NEURONAL APOPTOSIS.  
 CC -!- DISEASE: DEFECTS IN HD ARE THE CAUSE OF HUNTINGTON'S DISEASE, AN  
 CC AUTOSOMAL DOMINANT NEURODEGENERATIVE DISORDER CHARACTERIZED BY  
 CC INVOLUNTARY MOVEMENTS (CHOREA), GENERAL MOTOR IMPAIRMENT,  
 CC PSYCHIATRIC DISORDERS AND DEMENTIA. ONSET OF THE DISEASE OCCURS  
 CC USUALLY IN THE THIRD OR FOURTH DECADE OF LIFE AND SYMPTOMS  
 CC PROGRESSIVELY WORSEN LEADING TO DEATH IN 10 TO 20 YEARS. IT  
 CC AFFECTS 1 IN 10,000 INDIVIDUALS OF EUROPEAN ORIGIN. NEUROPATHOLOGY  
 CC OF HUNTINGTON'S DISEASE DISPLAYS A DISTINCTIVE PATTERN WITH LOSS  
 CC OF NEURONS, SPECIALLY IN THE CAUDATE AND PUTAMEN (STRIATUM).  
 CC -!- DATABASE: NAME=HotMolBase; NOTE=HD entry;  
 CC WWW="http://bioinformatics.weizmann.ac.il/hotmolebase/entries/hunti.htm".  
 CC -!- SIMILARITY: CONTAINS 10 HEAT REPEATS.  
 CC -!- SIMILARITY: BELONGS TO THE HUNTINGTIN FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration



```

FT REPEAT 1880 1891 REPEAT A.
FT REPEAT 1892 1902 REPEAT A (APPROXIMATE).
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH DOMAIN.
SQ SEQUENCE 3924 AA; 430337 MW; 0CC2249D CRC32;

Query Match 68.2%; Score 45; DB 1; Length 3924;
Best Local Similarity 40.0%; Pred. No. 1.34e+01;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 3839 DMPEIPPET 3848
QY :::|::|:
1 EEQEVPPD 10

RESULT 14
ID YKEL YEAST STANDARD; PRT; 224 AA.
AC P36095;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 26.2 KD PROTEIN IN SPC42-PTM1 INTERGENIC REGION.
GN YK1041W OR YK1254.
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
CC SACCCHAROMYCETACEAE; SACCCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE; 94205268.
RA PURNELLE B., TETTELIN H., VAN DYCK L., SKALA J., GOFFEAU A.;
RT "The sequence of a 17.5 kb DNA fragment on the left arm of yeast
RT chromosome XI identifies the protein kinase gene ELM1, the DNA
RT primase gene PRI2, a new gene encoding a putative histone and seven
RT new open reading frames.";
RL YEAST 9:1379-1384 (1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X71621; G656099; -
DR EMBL; Z28041; G486032; -
DR PIR; S37862; S37862.
DR HSSP; P03012; 2RSL.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 224 AA; 26242 MW; 3FEAA543 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 224;
Best Local Similarity 55.6%; Pred. No. 2.14e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 194 EEEQEIPDE 202
QY :::|::|:
1 EEQEVPPD 9

RESULT 15
ID DCUP ECOLI STANDARD; PRT; 354 AA.
AC P29680; P78135;
DT 01-FEB-1993 (REL. 25, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE UROPORPHYRINOGEN DECARBOXYLASE (EC 4.1.1.37) (UPD).
GN HEME.
OS ESCHERICHIA COLI.

```

---

```

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 94040783.
RA NISHIMURA K., INOKUCHI H.;
RT "Cloning and sequencing of the hemE gene encoding uroporphyrinogen
RT III decarboxylase (UPD) from Escherichia coli K-12.";
RL GENE 133:109-113 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 9408392.
RA BLATINER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes.";
RL NUCLEIC ACIDS RES. 21:5408-5417 (1993).
CC -!- CATALYTIC ACTIVITY: UROPORPHYRINOGEN III - COPROPORPHYRINOGEN +
CC 4 CO(2).
CC -!- PATHWAY: PORPHYRIN BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE UROPORPHYRINOGEN DECARBOXYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D12624; G216564; -
DR EMBL; U00006; G409791; -
DR EMBL; AE000473; G2367337; -
DR PIR; JS0708; JS0708.
DR ECGENE; EG11543; HEME.
DR PROSITE; PS00906; UROD_1; 1.
DR PROSITE; PS00907; UROD_2; 1.
DR PFAM; PF01208; URO-D; 1.
DR HSSP; P06132; URO.
KW LYASE; DECARBOXYLASE; PORPHYRIN BIOSYNTHESIS.
FT AILFSDI -> RSSRY (IN REF. 1).
FT D -> I (IN REF. 1).
FT LYEAGEGRFTSPV -> SSILKPEKVRVLRQI
FT (IN REF. 1).
FT CONFLICT 251 256
FT CONFLICT GGGOWL -> SATVA (IN REF. 1).
SQ SEQUENCE 354 AA; 39248 MW; A9638BF5 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 354;
Best Local Similarity 57.1%; Pred. No. 2.14e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 330 HDVPPE 336
QY :::|::|:
3 QOEVPDP 9

Search completed: Thu Oct 21 15:37:32 1999
Job time : 8 secs.

```

\*\*\*\*\*  
MPSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
Run on: Thu Oct 21 15:37:50 1999; MasPar time 4.30 Seconds  
Tabular output not generated.  
127.014 Million cell updates/sec

Title: >US-09-040-485-7  
Description: (1-10) from US09040485.pep  
Perfect Score: 66  
Sequence: 1 EEQEVPPDT 10

Scoring table: PAM 150  
Gap 15  
Searched: 179066 seqs, 54579741 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb1  
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified  
13:sp.vertebrate 14:sp.virus  
Statistics: Mean 21.069; Variance 28.505; scale 0.739

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	49	74.2	221 10 O65441 HYPOTHETICAL 24.8 KD P	6.97e+00
2	49	74.2	1406 4 O15082 KIAA0377	6.97e+00
3	49	74.2	1655 5 O24754 MASTERMIND	6.97e+00
4	48	72.7	698 4 Q93021 PUTATIVE TUMOR SUPPRES	1.09e+01
5	47	71.2	147 2 O66475 SINGLE STRANDED DNA-BI	1.71e+01
6	47	71.2	216 11 O08904 BRAIN X-LINKED GENE (B	1.71e+01
7	47	71.2	1120 5 Q20778 SIMILAR TO TRIPLE HELI	1.71e+01
8	46	69.7	68 13 Q91773 HEAT SHOCK PROTEIN (HS	2.65e+01
9	46	69.7	155 2 O85338 HYPOTHETICAL 17.3 KD P	2.65e+01
10	46	69.7	284 5 Q24013 G5-LIKE ORF'S PROTEIN	2.65e+01
11	46	69.7	360 10 Q40645 OSB28	2.65e+01
12	46	69.7	646 10 Q38970 CALMODULIN-DOMAIN PROT	2.65e+01
13	46	69.7	649 14 Q71093 TERMINAL PROTEIN	2.65e+01
14	46	69.7	704 10 O80896 PUTATIVE CONDENSIN PRO	2.65e+01
15	46	69.7	812 5 Q18177 C49H3.5 PROTEIN	2.65e+01
16	46	69.7	893 5 Q76417 MUTL HOMOLOG PMS2	2.65e+01
17	46	69.7	1320 14 Q06359 150 KD PROTEIN	2.65e+01
18	46	69.7	1828 14 Q89249 209 KDA READTHROUGH PR	2.65e+01
19	46	69.7	7962 4 Q10465 TITIN, SKELETAL MUSCLE	2.65e+01
20	45	68.2	370 5 O02271 F52F12.2 PROTEIN	4.09e+01

21	45	68.2	405	6	P79260	S-ANTIGEN (FRAGMENT)	4.09e+01
22	45	68.2	411	5	O26766	FLAGELLAR ANTIGEN (FRA	4.09e+01
23	45	68.2	1180	5	O18673	ZK270.2C PROTEIN	4.09e+01
24	45	68.2	1180	5	P91847	ZK270.2C-PROTEIN	4.09e+01
25	45	68.2	1463	2	O86919	AAS SURFACE PROTEIN	4.09e+01
26	44	66.7	382	14	O83924	ELB LARGE T PROTEIN	6.26e+01
27	44	66.7	411	3	O08412	ORF Y0R042W	6.26e+01
28	44	66.7	451	1	O06022	SOLUBLE TRANSDUCER PRO	6.26e+01
29	44	66.7	526	4	O75567	RAF RESPONSIVE ZINC FI	6.26e+01
30	44	66.7	584	5	O19672	F21C3.4 PROTEIN	6.26e+01
31	44	66.7	990	5	O46086	1-EVIDENCE-PREDICTED B	6.26e+01
32	44	66.7	1114	4	O75984	HYPOTHETICAL 127.6 KD	6.26e+01
33	44	66.7	1225	5	O17656	O5C9.3 PROTEIN	6.26e+01
34	44	66.7	1307	5	P91094	SIMILARITY TO DROSOPHI	6.26e+01
35	43	65.2	229	2	P96494	PUTATIVE GLYCINE DEHYD	9.53e+01
36	43	65.2	442	4	O15579	TRANSFORMING GROWTH FA	9.53e+01
37	43	65.2	664	4	O92541	MYELOBLAST KIAA0252 (F	9.53e+01
38	43	65.2	1012	11	O64028	EARLY DEVELOPMENT REGU	9.53e+01
39	43	65.2	1642	5	O62055	C09F9.2 PROTEIN	9.53e+01
40	43	65.2	2150	5	O17596	SDC-3 PROTEIN	9.53e+01
41	43	65.2	2559	5	O44113	PUTATIVE GUANINE NUCLE	9.53e+01
42	43	65.2	2912	4	O14054	COLLAGEN TYPE VII PREC	9.53e+01
43	43	65.2	2944	11	O63870	TYPE VII COLLAGEN	9.53e+01
44	43	65.2	3473	14	O83034	POLYPROTEIN	9.53e+01
45	43	65.2	4123	4	O75851	WUGSC-H_DJ0751H13.1 PR	9.53e+01

ALIGNMENTS

RESULT 1  
ID O65441 PRELIMINARY; PRT; 221 AA.  
AC O65441;  
DT 01-AUG-1998 (TREMREL. 07, CREATED)  
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 24.8 KD PROTEIN.  
GN F1C12.195.  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;  
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BEVAN M., TERRY N., ARDILES W., BUYSSHAERT C., DASSEVILLE R.,  
RA DE CLERCK R., DE KEYSER A., NEYT P., ROUZE P., VAN DEN DAELE H.,  
RA VILLAROE L., GIELEN J., VAN MONTAGU M., BANCROFT I., MEWES H.W.,  
RA MAYER K., SCHUELLER C.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU ARABIDOPSIS SEQUENCING PROJECT;  
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AL022224; E1283977;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 221 AA; 24834 MW; 7BEFAC97 CRC32;

Query Match 74.2%; Score 49; DB 10; Length 221;  
Best Local Similarity 60.0%; Pred. No. 6.97e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 14 EEQEEPPPS 23  
||| |||  
QY 1 EEQEVPPDT 10

RESULT 2	
ID O15082	PRELIMINARY; PRT; 1406 AA.
AC O15082;	
DT 01-JAN-1998 (TREMREL. 05, CREATED)	
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)	
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)	
DE KIAA0377.	
GN KIAA0377.	



OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE; 97349984.  
 RA NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,  
 RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA RES. 4:141-150(1997).  
 DR EMBL; AB002375; D1021673; -;  
 SQ SEQUENCE 1406 AA; 156319 MW; 68F9BC7E CRC32;

Query Match 74.2%; Score 49; DB 4; Length 1406;  
 Best Local Similarity 55.6%; Pred. No. 6.97e+00;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 44 DEDEVPPE 52  
 QY 1 EEQEVPPD 9

RESULT 3  
 ID Q24754 PRELIMINARY; PRT; 1655 AA.  
 AC Q24754;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE MASTERMIND.  
 OS DROSOPHILA VIRILIS (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
 OC DROSOPHILIDAE; DROSOPHILA.  
 RN [1]  
 RN SEQUENCE OF 260-762 FROM N.A.  
 RX MEDLINE; 91251140.  
 RA NEWFELD S.J., SMOLLER D.A., YEDVOBNICK B.;  
 RT "Interspecific comparison of the unusually repetitive Drosophila  
 RT locus mastermind.";  
 RL J. MOL. EVOL. 32:415-420(1991).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE; 94111143.  
 RA NEWFELD S.J., SCHMID A.T., YEDVOBNICK B.;  
 RT "Homopolymer length variation in the Drosophila gene mastermind.";  
 RL J. MOL. EVOL. 37:483-495(1993).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE; 94365848.  
 RA NEWFELD S.J., TACHIDA H., YEDVOBNICK B.;  
 RT "Drive-selection equilibrium: homopolymer evolution in the Drosophila  
 RT gene mastermind.";  
 RL J. MOL. EVOL. 38:637-641(1994).  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RA NEWFELD S.J.;  
 RL SUBMITTED (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; M92914; G157834; -;  
 DR FLYBASE; FBgn0013119; Dvir^man.  
 SQ SEQUENCE 1655 AA; 175048 MW; D521E17E CRC32;

Query Match 74.2%; Score 49; DB 5; Length 1655;  
 Best Local Similarity 55.6%; Pred. No. 6.97e+00;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 1084 QQQQVPPN 1092  
 QY 1 EEQEVPPD 9

RESULT 4  
 ID Q93021 PRELIMINARY; PRT; 698 AA.  
 AC Q93021;  
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE PUTATIVE TUMOR SUPPRESSOR.  
 GN LUCAL15.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA BENTLEY D., MAGGI L.;  
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RA WATERSTON R.;  
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U73168; G1613900; -;  
 DR PFAM; PF00076; rim; 1.  
 DR PFAM; PF00641; zf-RanBP; 1.  
 SQ SEQUENCE 698 AA; 78066 MW; 36A3242B CRC32;

Query Match 72.7%; Score 48; DB 4; Length 698;  
 Best Local Similarity 60.0%; Pred. No. 1.09e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 94 DSEQVPPGT 103  
 QY 1 EEQEVPPDT 10

RESULT 5  
 ID O66475 PRELIMINARY; PRT; 147 AA.  
 AC O66475;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE SINGLE STRANDED DNA-BINDING PROTEIN.  
 GN SSB.  
 OS AQUIFEX AEOLICUS.  
 OC BACTERIA; AQUIFICALES; AQUIFACACEAE; AQUIFEX.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE; 98196666.  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus.";  
 RL NATURE 392:353-358(1998).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AE000672; G2982816; -;  
 KW DNA-BINDING.  
 SQ SEQUENCE 147 AA; 17132 MW; BFA26F97 CRC32;

Query Match 71.2%; Score 47; DB 2; Length 147;  
 Best Local Similarity 75.0%; Pred. No. 1.71e+01;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 115 EEEVPP 122  
 QY 1 EEQEVPP 8

RESULT 6  
 ID O08904 PRELIMINARY: PRT: 216 AA.  
 AC O08904  
 DT 01-JUL-1997 (TREMREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE BRAIN X-LINKED GENE (BRX PROTEIN) (FRAGMENT).  
 GN BRX.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-BRAIN;  
 RA SIMLER M.C., HEARD E., ROUGEULLE C., CRUAUD C., WEISSENBACH J.,  
 RA AVNER P.;  
 RL MAMM. GENOME 0:0-0(0).  
 DR EMBL; Y11896; G311743; -.  
 DR MGD; MGI:109205; BRX.  
 FT NON\_TER 1  
 SQ SEQUENCE 216 AA; 24385 MW; 47AD381A CRC32;  
 Query Match 71.2%; Score 47; DB 11; Length 216;  
 Best Local Similarity 75.0%; Pred. No. 1.71e+01;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 76 EEEVVPP 83  
 QY 1 EEQEVPP 8  
 RESULT 7  
 ID Q20778 PRELIMINARY: PRT: 1120 AA.  
 AC Q20778  
 DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
 DE SIMILAR TO TRIPLE HELICAL REGION OF COLLAGENS.  
 GN F54D8.1.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;  
 OC RHABDITIDA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BOFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
 RL NATURE 0:0-0(0).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA BENTLEY D.;  
 RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U12966; G529221; -.  
 SQ SEQUENCE 1120 AA; 119368 MW; E0352B8E CRC32;  
 Query Match 71.2%; Score 47; DB 5; Length 1120;  
 Best Local Similarity 50.0%; Pred. No. 1.71e+01;  
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 127 EEEVVPAEA 136  
 QY 1 EEQEVPPDT 10  
 RESULT 8  
 ID Q91773 PRELIMINARY: PRT: 68 AA.  
 AC Q91773  
 DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE HEAT SHOCK PROTEIN (HSP30) (FRAGMENT).  
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;  
 OC MESOBATRACHIA; PIPOIDEA; PIPOIDAE; XENOPODINAE; XENOPUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 84221917.  
 RT BIENZ M.;  
 RL "Developmental control of the heat shock response in Xenopus.";  
 DR PROC. NATL. ACAD. SCI. U.S.A. 81:3138-3142(1984).  
 DR EMBL; K02305; G214267; -.  
 DR PFAM; PF00011; HSP20; 1.  
 KW HEAT SHOCK.  
 FT NON\_TER 1  
 FT NON\_TER 68  
 SQ SEQUENCE 68 AA; 7405 MW; 4D1E326C CRC32;  
 Query Match 69.7%; Score 46; DB 13; Length 68;  
 Best Local Similarity 55.6%; Pred. No. 2.65e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 48 DAQEIPPPDA 56  
 QY 2 EEQEVPPDT 10  
 RESULT 9  
 ID O85238 PRELIMINARY: PRT: 155 AA.  
 AC O85238  
 DT 01-NOV-1998 (TREMREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 17.3 KD PROTEIN.  
 OS YERSINIA ENTEROCOLITICA.  
 OG PLASMID PYV.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
 OC YERSINIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W22703;  
 RA IRIARTE M., KERBOURCH C., LAMBERMONT I., CORNELIS G.R.;  
 RL "Detail genetic map of the pyv plasmid of Y. enterocolitica O:9.";  
 RT SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AF054978; G341142; -.  
 KW HYPOTHETICAL PROTEIN; PLASMID.  
 SQ SEQUENCE 155 AA; 17309 MW; 65C5B6A CRC32;  
 Query Match 69.7%; Score 46; DB 2; Length 155;  
 Best Local Similarity 70.0%; Pred. No. 2.65e+01;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 131 ETQQLPPT 140  
 QY 1 EEQEVPPDT 10  
 RESULT 10  
 ID Q24013 PRELIMINARY: PRT: 264 AA.  
 AC Q24013  
 DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)



Db 391 EREEDVPP 398  
 QY 1 EEQEVPP 8  
 ||::|||

RESULT 14  
 ID O80896 PRELIMINARY; PRT; 704 AA.

AC O80896;  
 DT 01-NOV-1998 (TREMREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE PUTATIVE CONDENSIN PROTEIN.  
 GN T26B15.15.

OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
 OC CAPARALES; BRASSICACEAE; ARABIDOPSIS.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA ROUNSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,  
 RA BRANDON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,  
 RA SOMERVILLE C.R., VENTER J.C.;  
 RT "Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence."  
 RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AC004681; G3298547;  
 SQ SEQUENCE 704 AA; 79017 MW; 13BE2EF1 CRC32;

Query Match 69.7%; Score 46; DB 10; Length 704;  
 Best Local Similarity 50.0%; Pred. No. 2.65e+01;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 638 EEHQEPPES 647  
 QY 1 EEQEVPPDT 10  
 ||::|||

RESULT 15

ID Q18717 PRELIMINARY; PRT; 812 AA.

AC Q18717;  
 DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)  
 DE C49H3.5 PROTEIN.  
 GN C49H3.5.

OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
 RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
 RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
 RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
 RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
 RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A.,  
 RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLDMAN P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans."  
 RL NATURE 368:32-38(1994).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA WU X.;

RT "The sequence of C. elegans cosmid C49H3."  
 RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA WATERSTON R.;

RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U42436; G1123084;  
 SQ SEQUENCE 812 AA; 92341 MW; 132A227E CRC32;

Query Match 69.7%; Score 46; DB 5; Length 812;  
 Best Local Similarity 62.5%; Pred. No. 2.65e+01;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 518 EEQEQIPP 525  
 QY 1 EEQEVPP 8  
 ||::|||

Search completed: Thu Oct 21 15:38:42 1999  
 Job time : 52 secs.

\*\*\*\*\*  
M P S R C H  
\*\*\*\*\* (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:38:59 1999; MasPar time 3.32 Seconds  
Tabular output not generated. 64.003 Million cell updates/sec

Title: >US-09-040-485-8  
Description: (1-10) from US09040485.pap  
Perfect Score: 67  
Sequence: 1 DGPTGEPQQE 10  
Scoring table: PAM 150  
Gap 15

Searched: 170751 seqs, 21266508 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 14.838; Variance 44.503; scale 0.333  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Pred. No.
1	47	70.1	749 21	W14053	8.12e+01
2	47	70.1	1122 21	W14052	8.12e+01
3	47	70.1	1389 21	W14051	8.12e+01
4	47	70.1	3567 8	R4431	8.12e+01
5	45	67.2	1341 13	R71701	1.35e+02
6	45	67.2	1418 13	R59751	1.35e+02
7	45	67.2	1418 13	R71703	1.35e+02
8	45	67.2	1487 34	W61561	1.35e+02
9	44	65.7	243 10	E1A 243	1.73e+02
10	44	65.7	745 20	W03560	1.73e+02
11	44	65.7	745 20	R76110	1.73e+02
12	43	64.2	943 28	W43039	2.23e+02
13	43	64.2	3080 1	P93285	2.23e+02
14	43	64.2	3210 1	P81771	2.23e+02
15	42	62.7	17 2	R06684	2.85e+02
16	42	62.7	19 5	R25268	2.85e+02

17	42	62.7	25 6	R31213	HPV-16 E7 peptide.	2.85e+02
18	42	62.7	98 4	R22767	HPV E7 peptide.	2.85e+02
19	42	62.7	98 29	W46886	Amino acid sequence o	2.85e+02
20	42	62.7	98 8	R42361	Human papillomavirus	2.85e+02
21	42	62.7	172 19	R97562	Human papilloma virus	2.85e+02
22	42	62.7	181 17	R94588	Rat R33 platelet deri	2.85e+02
23	42	62.7	181 17	R94589	Human R33 platelet de	2.85e+02
24	42	62.7	200 15	R76871	Talin C-terminal pept	2.85e+02
25	42	62.7	253 38	W81586	CTLA4/E7 fusion prote	2.85e+02
26	42	62.7	253 38	W87562	CTLA4-E7 fusion prote	2.85e+02
27	42	62.7	266 19	R97561	Human papilloma virus	2.85e+02
28	42	62.7	324 24	W24971	Rat amelin-2 protein.	2.85e+02
29	42	62.7	407 24	W24970	Rat amelin-1 protein.	2.85e+02
30	42	62.7	652 36	W72233	HSV-2 strain SB5 Cont	2.85e+02
31	42	62.7	718 36	W72072	HSV-2 strain SB5 Cont	2.85e+02
32	42	62.7	725 36	W72234	HSV-2 strain SB5 Cont	2.85e+02
33	42	62.7	754 36	W72235	HSV-2 strain SB5 Cont	2.85e+02
34	42	62.7	772 36	W72236	HSV-2 strain SB5 Cont	2.85e+02
35	42	62.7	791 36	W72237	HSV-2 strain SB5 Cont	2.85e+02
36	42	62.7	791 36	W72238	HSV-2 strain SB5 Cont	2.85e+02
37	42	62.7	819 36	W72240	HSV-2 strain SB5 Cont	2.85e+02
38	42	62.7	826 36	W72241	HSV-2 strain SB5 Cont	2.85e+02
39	42	62.7	1442 14	R79480	Rat type II collagen.	2.85e+02
40	42	62.7	2237 14	R71006	Human neuronal calciu	2.85e+02
41	42	62.7	2237 6	R33550	Sequence of the alpha	2.85e+02
42	42	62.7	2237 33	W63142	Human calcium channel	2.85e+02
43	42	62.7	2337 32	W37878	Human calcium channel	2.85e+02
44	42	62.7	2339 33	W63141	Human calcium channel	2.85e+02
45	42	62.7	2339 14	R71005	Human neuronal calciu	2.85e+02

ALIGNMENTS

RESULT 1  
ID W14053 standard; Protein; 749 AA.  
AC W14053;  
DT 28-MAY-1997 (first entry)  
DE TIM01 mutant protein.  
KW TIM; timeless protein; nuclear translocation protein; circadian rhythm;  
KW sleep-wake cycle; light sensitive; environmental cycle; nuclear protein;  
KW Drosophila period gene; transcription factor; PAS domain; depression;  
KW narcolepsy; PER; jet lag; NTP; therapy.  
OS Drosophila melanogaster.  
FH Key Location/Qualifiers  
FT region 351..380  
FT /note= "acidic region"  
FT region 540..553  
FT /note= "basic region, possible nuclear localisation  
FT signal"  
PN WO9629406-A2.  
PD 26-SEP-1996.  
PE 20-MAR-1996; U03830.  
PF 20-MAR-1995; US-408518.  
PR 20-MAY-1995; US-442214.  
PR 16-MAY-1995; US-442214.  
PR 02-NOV-1995; US-552354.  
PA (UYRQ ) UNIV PENNSYLVANIA.  
PA (UYRQ ) UNIV ROCKEFELLER.  
PI Myers MP, Price JL, Sehgal A, Vossell LB, Young MW;  
DR WPI: 96-443182/44.  
DR N-PSDB; T60336.  
PT Nuclear translocation protein which binds to protein involved in  
PT circadian rhythms - used to develop prods. for diagnosis, prevention  
PT or treatment of disorders associated with circadian rhythms, e.g.  
PT jetlag and narcolepsy  
PS Example 9; : 131pp; English.  
CC This sequence represents the 01 mutant of the Drosophila melanogaster  
CC "timeless" (TIM) protein. The TIM01 mutant is an arrhythmic mutation. TIM  
CC is a nuclear translocation protein (NTP) of the invention. The NTPs of  
CC the invention have specific binding activity to a protein involved in  
CC circadian rhythms, and cyclic transcription patterns related to the  
CC sleep-wake cycle. The NTP is also preferably light sensitive, and has a  
CC stabilising effect on the circadian rhythm protein. The NTP also has the  
CC ability to aid the process of circadian rhythm entrainment to

CC environmental cycles of light. TIM has specific binding activity for the  
 CC Drosophila period (PER) gene. PER is a nuclear protein which has homology  
 CC to the family of transcription factors containing the PAS domain, and is  
 CC a protein involved in circadian rhythms, but the biological function of  
 CC PER is unknown. The amount of PER fluctuates with a circadian rhythm,  
 CC and the protein is also phosphorylated with a circadian rhythm. The NTPs  
 CC of the invention, their fragments, agonists, mimics and antagonists are  
 CC useful for preventing and/or treating disorders of a circadian rhythm  
 CC such as depression, narcolepsy or jet lag. The products can also be used  
 CC for detecting and/or measuring conditions so as to classify groups of  
 CC individuals with the disorders.  
 SO Sequence 749 AA;

Query Match 70.1%; Score 47; DB 21; Length 749;  
 Best Local Similarity 60.0%; Pred. No. 8.12e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 dpggkpkqh 519  
 ||| | ||:  
 QY 1 DPGTGEPOQE 10

RESULT 2  
 ID W14052 standard; Protein; 1122 AA.

AC W14052;  
 DT 28-MAY-1997 (first entry)

DE TIM protein.

KW TIM; timeless protein; nuclear translocation protein; circadian rhythm;  
 KW sleep-wake cycle; light sensitive; environmental cycle; nuclear protein;  
 KW Drosophila period gene; transcription factor; PAS domain; depression;  
 KW narcolepsy; PER; jet lag; NTP; therapy.

OS Drosophila melanogaster.

FH Key Location/Qualifiers

FT region 351..380

FT /note= "acidic region"

FT region 540..553

FT /note= "basic region, possible nuclear localisation  
 signal"

FT W09629406-A2.

PD 26-SEP-1996.

PF 20-MAR-1996; U03830.

PR 20-MAR-1995; US-408518.

PR 16-MAY-1995; US-442214.

PR 02-NOV-1995; US-552354.

PA (UYPE-) UNIV PENNSYLVANIA.

PI (UYRQ) UNIV ROCKEFELLER.

PI Myers MP, Price JL, Sehgal A, Vossell LB, Young MW;

DR WPI: 96-443182/44.

DR N-PSDB; T60333, T60335.

DR Nuclear translocation protein which binds to protein involved in

PT circadian rhythms - used to develop prods. for diagnosis, prevention

PT or treatment of disorders associated with circadian rhythms, e.g.

PT jetlag and narcolepsy

PS Claim 3; Fig 11; 131pp; English.

CC This sequence represents the Drosophila melanogaster "timeless" (TIM)

CC protein. TIM is a nuclear translocation protein (NTP) of the invention.

CC The NTPs of the invention have specific binding activity to a protein

CC involved in circadian rhythms, and cyclic transcription patterns related

CC to the sleep-wake cycle. The NTP is also preferably light sensitive, and

CC has a stabilising effect on the circadian rhythm protein. The NTP also

CC has the ability to aid the process of circadian rhythm entrainment to

CC environmental cycles of light. TIM has specific binding activity for the

CC Drosophila period (PER) gene. PER is a nuclear protein which has homology

CC to the family of transcription factors containing the PAS domain, and is

CC a protein involved in circadian rhythms, but the biological function of

CC PER is unknown. The amount of PER fluctuates with a circadian rhythm,

CC and the protein is also phosphorylated with a circadian rhythm. The NTPs

CC of the invention, their fragments, agonists, mimics and antagonists are

CC useful for preventing and/or treating disorders of a circadian rhythm

CC such as depression, narcolepsy or jet lag. The products can also be used

CC for detecting and/or measuring conditions so as to classify groups of

CC individuals with the disorders.

SO Sequence 1122 AA;

Query Match 70.1%; Score 47; DB 21; Length 1122;  
 Best Local Similarity 60.0%; Pred. No. 8.12e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 dpggkpkqh 519  
 ||| | ||:  
 QY 1 DPGTGEPOQE 10

RESULT 3

ID W14051 standard; Protein; 1389 AA.

AC W14051;

DT 28-MAY-1997 (first entry)

DE TIM protein splice variant.

KW TIM; timeless protein; nuclear translocation protein; circadian rhythm;

KW sleep-wake cycle; light sensitive; environmental cycle; nuclear protein;

KW Drosophila period gene; transcription factor; PAS domain; depression;

KW narcolepsy; PER; jet lag; NTP; therapy.

OS Drosophila melanogaster.

FH Key Location/Qualifiers

FT region 351..380

FT /note= "acidic region"

FT region 540..553

FT /note= "basic region, possible nuclear localisation

FT signal"

FT W09629406-A2.

PD 26-SEP-1996.

PF 20-MAR-1996; U03830.

PR 20-MAR-1995; US-408518.

PR 16-MAY-1995; US-442214.

PR 02-NOV-1995; US-552354.

PA (UYPE-) UNIV PENNSYLVANIA.

PI (UYRQ) UNIV ROCKEFELLER.

PI Myers MP, Price JL, Sehgal A, Vossell LB, Young MW;

DR WPI: 96-443182/44.

DR N-PSDB; T60334.

DR Nuclear translocation protein which binds to protein involved in

PT circadian rhythms - used to develop prods. for diagnosis, prevention

PT or treatment of disorders associated with circadian rhythms, e.g.

PT jetlag and narcolepsy

PS Claim 3; Fig 11; 131pp; English.

CC This sequence represents the splice variant of the Drosophila

CC melanogaster "timeless" (TIM) protein. TIM is a nuclear translocation

CC protein (NTP) of the invention. The NTPs of the invention have specific

CC binding activity to a protein involved in circadian rhythms, and cyclic

CC transcription patterns related to the sleep-wake cycle. The NTP is also

CC preferably light sensitive, and has a stabilising effect on the circadian

CC rhythm protein. The NTP also has the ability to aid the process of

CC circadian rhythm entrainment to environmental cycles of light. TIM has

CC specific binding activity for the Drosophila period (PER) gene. PER is a

CC nuclear protein which has homology to the family of transcription factors

CC containing the PAS domain, and is a protein involved in circadian

CC rhythms, but the biological function of PER is unknown. The amount of

CC PER fluctuates with a circadian rhythm, and the protein is also

CC phosphorylated with a circadian rhythm. The NTPs of the invention, their

CC fragments, agonists, mimics and antagonists are useful for preventing

CC and/or treating disorders of a circadian rhythm such as depression,

CC narcolepsy or jet lag. The products can also be used for detecting and/or

CC measuring conditions so as to classify groups of individuals with the

CC disorders.

SO Sequence 1389 AA;

Query Match 70.1%; Score 47; DB 21; Length 1389;  
 Best Local Similarity 60.0%; Pred. No. 8.12e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 dpggkpkqh 519  
 ||| | ||:  
 QY 1 DPGTGEPOQE 10

RESULT 4

ID R44431 standard; Protein; 3567 AA.  
AC R44431;  
DT 22-DEC-1993 (first entry)  
DE eryA region polypeptide module #2.  
KW Saccharopolyspora erythraea; eryA; biosynthesis; polypeptide; module;  
KW erythromycin; condensation; elongation; acyl chain growth;  
KW gene replacement.  
OS Saccharopolyspora erythraea.  
PN WO9313663-A.  
PD 22-JUL-1993.  
PF 17-JAN-1992; U00427.  
PR 17-JAN-1992; WO-U00427.  
PA (ABBO) ABBOTT LAB.  
PI Donadio S, Katz L, McAlpine JB;  
DR WPI: 93-242804/30.  
DR N-PSDB: Q46806.  
PT Biosynthesis of specific polypeptide analogues esp. erythromycin  
PT cpds. - by introducing altered biosynthetic gene-contg. DNA into  
PT microorganisms  
PS Disclosure; Fig 2; 133pp; English.  
CC The sequences given in R44430-32 are encoded by the eryA fragment of  
CC the Saccharopolyspora erythraea genome. These polypeptides are  
CC involved in the biosynthesis of the polypeptide segment of erythromycin.  
CC eryA is organised in modules and each module takes care of one  
CC condensation step. The precise succession of elongation steps is  
CC dictated by the genetic order of the modules. The DNA encoding  
CC these polypeptides may be specifically altered such that novel  
CC polypeptide molecules of desired structure are produced. Three types  
CC of alteration may be produced; those inactivating a single function in  
CC a module which does not arrest acyl chain growth; those inactivating a  
CC single function in a module which does affect chain growth; and those  
CC affecting an entire module. The mutations may be introduced by gene  
CC replacement.  
SQ Sequence 3567 AA;

Query Match 70.1%; Score 47; DB 8; Length 3567;  
Best Local Similarity 87.5%; Pred. No. 8.12e+01;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1376 gptgepaq 1383  
|||||  
QY 2 GPTGEPOQ 9

RESULT 5  
ID R71701 standard; protein; 1341 AA.  
AC R71701;  
DE Collagen alpha 1 (I) chain precursor.  
KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;  
KW disorder; osteoporosis; metastatic progression; Paget's disease;  
KW hyperthyroidism; bone; resorption; rheumatoid arthritis;  
KW osteoarthritis; vasculitis syndrome.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT misc\_difference 2028 /note= "Unidentified amino acid."

Query Match 70.1%; Score 47; DB 8; Length 3567;  
Best Local Similarity 87.5%; Pred. No. 8.12e+01;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1376 gptgepaq 1383  
|||||  
QY 2 GPTGEPOQ 9

RESULT 5  
ID R71701 standard; protein; 1341 AA.  
AC R71701;  
DE Collagen alpha 1 (I) chain precursor.  
KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;  
KW disorder; osteoporosis; metastatic progression; Paget's disease;  
KW hyperthyroidism; bone; resorption; rheumatoid arthritis;  
KW osteoarthritis; vasculitis syndrome.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT misc\_difference 2028 /note= "Unidentified amino acid."

Query Match 70.1%; Score 47; DB 8; Length 3567;  
Best Local Similarity 87.5%; Pred. No. 8.12e+01;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1376 gptgepaq 1383  
|||||  
QY 2 GPTGEPOQ 9

RESULT 5  
ID R71701 standard; protein; 1341 AA.  
AC R71701;  
DE Collagen alpha 1 (I) chain precursor.  
KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;  
KW disorder; osteoporosis; metastatic progression; Paget's disease;  
KW hyperthyroidism; bone; resorption; rheumatoid arthritis;  
KW osteoarthritis; vasculitis syndrome.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT misc\_difference 2028 /note= "Unidentified amino acid."

CC collagen metabolism (degradation of type I collagen may indicate  
CC osteoporosis, metastatic progression, Paget's disease,  
CC hyperthyroidism or other conditions involving excessive bone  
CC resorption; degradation of type II collagen may indicate rheumatoid  
CC arthritis or osteoarthritis; and of type III collagen, vacuolitis  
CC syndrome). The method can also be used to assess the toxicity of a  
CC compound and to test drugs for their effect on collagen metabolism.  
SQ Sequence 1341 AA;

Query Match 67.2%; Score 45; DB 13; Length 1341;  
Best Local Similarity 44.4%; Pred. No. 1.35e+02;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 202 zgpgpbgz 210  
|||||  
QY 1 DGPGEPOQ 9

## RESULT 6

ID R59751 standard; protein; 1418 AA.  
AC R59751;  
DT 14-FEB-1995 (first entry)  
DE Type II collagen.  
KW Collagen; triple helix; articular cartilage; collagenase;  
KW degradation; monoclonal antibody; epitope; matrix;  
KW metalloproteinase.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT peptide 1..24 /label= Signal peptide.  
PN WO9414070-A.  
PD 23-JUN-1994.  
PF 06-DEC-1993; CA0522.  
PR 04-DEC-1992; US-984123.  
PA (SHRI-) SHRINERS HOSPITALS FOR CRIPPLED CHILDREN.  
PI Hollander AP, Poole AR;  
DR WPI: 94-234222/28.  
PT Determn. of cartilage degradation - using a monoclonal antibody  
PT to measure the amt. of unwound collagen or fragments in samples  
PS Disclosure; Figure 1; 119pp; English.  
CC Type II collagen constitutes the bulk of the fibrillar backbone of  
CC cartilage matrix. It is composed of a tightly wound triple helix  
CC which can only be cleaved by the metalloproteinase collagenase to  
CC produce 3/4 and 1/4 length alpha chain fragments. The destruction of  
CC articular cartilage is due, in part, to the degradation of collagen.  
CC Incapable of maintaining its helical structure at physiological  
CC temperatures, collagenase-cleaved collagens unwind and become  
CC susceptible to further degradation by other proteinases. By  
CC producing monoclonal antibodies directed against epitopes which are  
CC only revealed when collagen is unwound, the antibodies provide a  
CC means of determining the degradation of cartilage in a biological  
CC sample. The antibodies do not bind to native helical collagen.  
CC Epitopes used in the production of such antibodies are described in  
CC R59749, R59750 and R67742.  
SQ Sequence 1418 AA;

Query Match 67.2%; Score 45; DB 11; Length 1418;  
Best Local Similarity 66.7%; Pred. No. 1.35e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 930 gpgpgpqq 938  
|||||  
QY 2 GPTGEPOQ 10

## RESULT 7

ID R71703 standard; protein; 1418 AA.  
AC R71703;  
DT 17-OCT-1995 (first entry)  
DE Collagen alpha 1 (II) chain precursor.  
KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;  
KW disorder; osteoporosis; metastatic progression; Paget's disease;  
KW hyperthyroidism; bone; resorption; rheumatoid arthritis;

KW osteoarthritis; vasculitis syndrome.  
 OS Homo sapiens.  
 PN W09508115-A.  
 PD 23-MAR-1995.  
 PF 19-SEP-1994; DK0348.  
 PR 17-SEP-1993; DK-001040.  
 PA (OSTE-) OSTEOMETER AS.  
 PI Bonde M, Qvist P;  
 DR WPI; 95-131456/17.  
 PT Assaying collagen fragments in body fluid by immunoassay - using  
 PT antibodies raised against synthetic peptide(s) contg. potential  
 PT crosslinking sites, to diagnose and monitor disorders of collagen  
 PT metabolism, e.g. osteoporosis.  
 PS Disclosure (Appendix A); Page 53; 87pp; English.  
 CC Determination of collagen fragments in body fluids can be achieved  
 CC by immunoassay using antibodies directed against synthetic peptides  
 CC derived from collagen which contain sites of potential crosslinking.  
 CC The method is used to diagnose and monitor treatment of disorders of  
 CC collagen metabolism (degradation of type I collagen may indicate  
 CC osteoporosis, metastatic progression, Paget's disease,  
 CC hyperthyroidism or other conditions involving excessive bone  
 CC resorption; degradation of type II collagen may indicate rheumatoid  
 CC arthritis or osteoarthritis; and of type III collagen, vaculitis  
 CC syndrome). The method can also be used to assess the toxicity of a  
 CC compound and to test drugs for their effect on collagen metabolism.  
 SQ Sequence 1418 AA;  
  
 Query Match 67.2%; Score 45; DB 13; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 1.35e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
 Db 930 gspgpggg 938  
 ||:|||||  
 QY 2 GPTGEPQOE 10  
  
 RESULT 8  
 ID W61562 standard; protein; 1487 AA.  
 AC W61562;  
 DE Human type II collagen alpha-chain protein.  
 KW Type II collagen alpha-chain; human; immunoassay; antibody; C-Iifree;  
 KW epitope; cartilage; enzyme-linked immunosorbent assay; ELISA; therapy;  
 KW diagnosis; arthritis; growth disorder; prognosis; drug screening;  
 KW anti-arthritis agent; matrix metalloproteinase inhibitor.  
 OS Homo sapiens.  
 PN W09835235-A1.  
 PD 13-AUG-1998.  
 PF 30-JAN-1998.  
 PR 08-FEB-1997; GB-002252.  
 PA (YISH-) UNIV SHEFFIELD.  
 PI Croucher LJ, Hollander AP;  
 DR WPI; 98-447376/38.  
 PT Immunoassay kit containing two antibodies recognising coupled  
 PT epitope(s) on collagen fragments - and new antibodies, for  
 PT diagnosing arthritis etc., also prognosis and screening for  
 PT anti-arthritis agents or inhibitors of matrix metallo-protease  
 PS Disclosure; Fig 2; 57pp; English.  
 CC This sequence represents the human type II collagen alpha-chain which  
 CC is used in a method to produce an immunoassay kit comprising of two  
 CC antibodies (Ab1 and Ab2), mono- or poly-clonal, or their fragments,  
 CC that bind to two C-Iifree coupled epitopes (C-Iifree indicates any type  
 CC II collagen fragment that is released from degraded cartilage). The kits  
 CC are designed for sandwich immunoassays, specifically enzyme-linked  
 CC immunosorbent assay (ELISA), and C-Iifree is systemic (present in urine,  
 CC serum or synovial fluid). The kits are used for therapy, diagnosis (e.g.  
 CC rheumatoid arthritis and osteoarthritis, or monitoring progression of  
 CC growth hormone) and for drug screening (to identify, and assess efficacy  
 CC of, anti-arthritis agents and matrix metalloproteinase inhibitors).  
 CC C-Iifree, derived from the N-terminus of the alpha 1 chain, have  
 CC increased resistance to proteolysis, so can accumulate in vivo to a

CC concentration that allows accurate measurement by immunoassay.  
 SQ Sequence 1487 AA;  
  
 Query Match 67.2%; Score 45; DB 34; Length 1487;  
 Best Local Similarity 66.7%; Pred. No. 1.35e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
 Db 999 gspgpggg 1007  
 ||:|||||  
 QY 2 GPTGEPQOE 10  
  
 RESULT 9  
 ID R52601 standard; Protein; 243 AA.  
 AC R52601;  
 DE E1A 243 amino acid protein.  
 KW Alternative splicing; E1A; adenovirus; early gene; oncogenes;  
 KW transcription regulation protein; oncogenic transformation; rodent;  
 KW phenotype; human; tumour cell; contact inhibition; differentiation;  
 KW adenovirus 5; anchorage-independant growth; tumorigenic potential;  
 KW reorganisation; flat morphology; tumour suppressor.  
 OS Adenovirus.  
 PN W09409160-A.  
 PD 28-APR-1994.  
 PF 13-OCT-1993; U09774.  
 PR 13-OCT-1992; US-960112.  
 PA (LJOL-) LA JOLLA CANCER RES FOUND.  
 PI Frisch SM;  
 DR WPI; 94-151344/18.  
 DR N-PSDB; Q62526.  
 PT Restoring contact inhibition to hyperproliferative cells - by  
 PT introducing nucleic acid encoding E1A peptide, also promoting  
 PT differentiation, used for treating malignancies  
 PS Disclosure; Page 20-22; 41pp; English.  
 CC This sequence is encoded by the product of alternative splicing of the  
 CC E1A RNA such that the information contained within this sequence is a  
 CC subset of the larger E1A protein, see also R52602. E1A is an adeno-  
 CC virus early gene and produces two products. The 243 and 289 residue  
 CC proteins are both transcriptional regulation proteins which facilitate  
 CC the oncogenic transformation of certain rodent cells by other  
 CC oncogenes. The adenovirus E1A gene unexpectedly influences the  
 CC phenotype of human tumour cells so as to restore their contact  
 CC inhibitory properties and promote differentiation. Stable-expression  
 CC of the adenovirus 5 E1A gene reduces anchorage-independant growth and  
 CC tumorigenic potential, promotes reorganisation, induces flat  
 CC morphology, and restores contact inhibition in human tumour cell  
 CC lines. Therefore E1A acts as a tumour suppressor gene in this human  
 CC context.  
 SQ Sequence 243 AA;  
  
 Query Match 65.7%; Score 44; DB 10; Length 243;  
 Best Local Similarity 50.0%; Pred. No. 1.73e+02;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
 Db 138 egpvsepepe 147  
 ||:|||||  
 QY 1 DGPTGEPQOE 10  
  
 RESULT 10  
 ID W03560 standard; Protein; 745 AA.  
 AC W03560;  
 DE 26-FEB-1997 (first entry)  
 DE Human adrenoleucodystrophy wild-type protein ALDP.  
 KW Adrenoleucodystrophy; gene therapy; retroviral vector M48;  
 KW adrenoleucopathy; membrane protein; long chain fatty acid oxidation.  
 OS Homo sapiens.  
 PN W09621733-A2.  
 PD 18-JUL-1996.  
 PF 12-JAN-1996; F00059.  
 PR 13-JAN-1995; FR-000376.  
 PA (INRM) INST NAT SANTE & RECH MEDICALE.



PI Aubourg P, Mandel JL, Mosser J, Sarde CO;  
 DR WPI: 96-342286/34.  
 DR N-PSDB; T39335.  
 PT Recombinant viral vector contg. DNA for correcting  
 PT adrenoleuco-dystrophy - and immunologically tolerable cells contg.  
 PT this vector, useful in gene therapy  
 PS Example: Fig 7; 36pp; French.  
 CC A 2.43 kb SpeI-EcoRI fragment of human wild-type ALD  
 CC (adrenoleucodystrophy) cDNA was inserted into retroviral vector M48  
 CC so that it was under control of the mouse phosphoglycerate kinase  
 CC (PGK) promoter. The resulting vector was co-transfected with vector  
 CC PGK-neomycin into amphotropic packaging cell line psiCRIP. Neomycin  
 CC (G418)-resistant transduced cells which were also positive with  
 CC anti-ALD antibodies were incubated with skin fibroblasts of an ALD  
 CC patient. After two rounds of infection, about 70% of cells were  
 CC expressing normal ALD protein (ALDP) and integration of M48-ALD was  
 CC confirmed by Southern blotting. The present sequence is that of  
 CC ALDP.  
 SQ Sequence 745 AA;

Query Match 65.7%; Score 44; DB 20; Length 745;  
 Best Local Similarity 60.0%; Pred. No. 1.73e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 gapageptqg 56  
 QY 1 DGPTGEPQGE 10

## RESULT 11

ID R76110 standard; Protein; 745 AA.  
 AC R76110.  
 DT 16-NOV-1995 (first entry)  
 DE Human ALD.  
 KW ALD gene; adrenoleukodystrophy; adrenomyeloneuropathy; AMN;  
 OS diagnosis; gene therapy.  
 OS Homo sapiens.  
 PN CA2108606-A.  
 PD 16-APR-1995.  
 PF 15-OCT-1993; 108606.  
 PR 15-OCT-1993; CA-108606.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PI Aubourg P, Mandel J, Mosser J, Sarde C;  
 DR WPI: 95-215721/29.  
 DR N-PSDB; Q94048.

PT New nucleic acid responsible for adreno-leuco-dystrophy - related  
 PT probes, proteins and antibodies, useful for diagnosis and treatment  
 PS Claim 16; Fig.2A-B; 39pp; English.  
 CC Probes corresp. to breakpoints in the red pigment gene of an AMN  
 CC patient were used to isolate clones from an Xq28 cosmid library.  
 CC Following hybridizations and nested PCR, a HeLa cell cDNA library  
 CC was screened, and a complete ALD cDNA fragment (Q94048) encoding the  
 CC protein given in R94048 was obtd. Transformation of hematopoietic  
 CC cells with ALD-encoding sequences in vivo will allow therapy of ALD  
 CC or AMN.

SQ Sequence 745 AA;

Query Match 65.7%; Score 44; DB 13; Length 745;  
 Best Local Similarity 60.0%; Pred. No. 1.73e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 gapageptqg 56  
 QY 1 DGPTGEPQGE 10

## RESULT 12

ID W43039 standard; Protein; 943 AA.  
 AC W43039;  
 DT 12-MAY-1998 (first entry)  
 DE A NADPH oxidase derived from Arabidopsis thaliana.  
 KW NADPH oxidase; development; agricultural chemical; growth control;  
 KW transformation; plant; evaluation.

OS Arabidopsis thaliana.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 161 /note= "not specified"  
 FT Misc\_difference 165 /note= "not specified"  
 FT  
 PN J10033176-A.  
 PD 10-FEB-1998.  
 PF 23-JUL-1996; 193220.  
 PR 23-JUL-1996; JP-193220.  
 PA (SUMO ) SUMITOMO CHEM CO LTD.  
 DR WPI: 98-172095/16.  
 DR N-PSDB; V05045.  
 PT NADPH oxidase derived from plants - useful for control of growth of  
 PT transformed plants and for evaluation of agricultural chemicals  
 PS Claim 1; Pages 7-9; ilpp; Japanese.  
 CC The presents sequence represents a NADPH oxidase. The cDNA encoding  
 CC this protein sequence was isolated from a cDNA library of Arabidopsis  
 CC thaliana. The NADPH DNA and protein can be used in the development of  
 CC agricultural chemicals. They allow for the control of growth of  
 CC transformed plants for smooth evaluation of agricultural chemicals.  
 SQ Sequence 943 AA;

Query Match 64.2%; Score 43; DB 28; Length 943;  
 Best Local Similarity 60.0%; Pred. No. 2.23e+02;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 732 dgpygapaqg 741  
 QY 1 DGPTGEPQGE 10

## RESULT 13

ID P93285 standard; protein; 3080 AA.  
 AC P93285;  
 DT 06-APR-1990 (first entry)  
 DE Sequence of clone HIV-2 SBL/ISY.  
 KW HIV-2; proviral clone HIV-2 SBL/ISY.  
 OS Human immunodeficiency virus 2.  
 PN US7331212-A.  
 PD 29-AUG-1989.  
 PF 31-MAR-1989; 331212.  
 PR 31-MAR-1989; US-331212.  
 PA (USSH) US Dept. Health and Human Services.  
 PI Franchini G, Wong-Staal F, Gallo R;  
 DR WPI: 89-339698/46.  
 DR N-PSDB; N92119.

PT Complete human immunodeficiency type 2 proviral clone - used to generate  
 PT animal model for function studies of HIV genes in vivo.  
 PS Disclosure; Fig. 5; 43pp; English.  
 CC The protein is encoded by the third reading frame of HIV-2 SBL/ISY, a  
 CC proviral clone of HIV-2.  
 SQ Sequence 3080 AA;

Query Match 64.2%; Score 43; DB 1; Length 3080;  
 Best Local Similarity 75.0%; Pred. No. 2.23e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 2799 phgepqgq 2806  
 QY 3 PTGEPQGE 10

## RESULT 14

ID P81771 standard; protein; 3210 AA.  
 AC P81771;  
 DT 07-NOV-1990 (first entry)  
 DE Deduced sequence encoded by bottom reading frame of cDNA clone  
 DE HIV-2 SBL/ISY of HIV related retrovirus strain  
 KW HIV vaccine; HIV strain SBL-6669-85.  
 OS Human immunodeficiency virus.  
 PN W08808449-A.  
 PD 03-NOV-1988.

QY 1 DGPTGPEQOE 10

Search completed: Thu Oct 21 15:39:19 1999  
Job time : 20 secs.

PF 28-APR-1988; SE0218.  
PR 28-APR-1987; SE-001765.  
PA (SLS-) SBL Statens Bacteri.  
PI Albert J, Biberfeld G, Fenyo EM, Norrby E;  
DR WPI; 88-322769/45.  
DR NP-PSDB; n80890.  
PT HIV related human retro-virus strain -  
PT used for obtaining antigens for assays and vaccines and for  
PT prodn. of antibodies for assays  
PS Claim 9; Fig 4; 28pp; English.  
CC Synthetically produced proteins and peptides, characterised in that the  
CC AA sequence is derived from the primary nucleotide sequence of  
CC HIV-2 SBL/ISV or a part thereof, or a degenerate thereof are claimed.  
CC HIV-2 SBL/ISV represents the complete genome of the virus SBL-6699  
CC (-SBL-6689-85). The proviral DNA was obtd. from a genomic library  
CC constructed from DNA of HUT-78 cells infected with SBL-6669-85 using  
CC the lambda-phage vector EMBL-3. SBL-6669-85 was isolated from lymphocytes  
CC of a West African woman. Protection is requested for the entire genome  
CC disclosed in n80890 and for parts thereof, and corresp. to various genes  
CC such as the gag gene (corresp. to nucleotides 547 to 2106), the pol gene  
CC (nucleotides 1827-4931) and the env gene (nucleotides 6144 to 8682), the  
CC corresp. AA sequences and parts thereof and various products derived  
CC therefrom, or use thereof, such as clones prepd. by recombinant vector  
CC method, HIV test devices and methods. X corresponds to the translation of  
CC a stop codon.  
SQ Sequence 3210 AA;

Query Match 64.2%; Score 43; DB 1; Length 3210;  
Best Local Similarity 75.0%; Pred. NO. 2.23e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 2906 phgqpqq 2913

QY 3 PTGEPQOE 10

RESULT 15

ID R06684 standard; protein; 17 AA.  
AC R06684;  
DT 11-JAN-1991 (first entry)  
DE Human papilloma virus 16 E7-protein DNA sequence (V).  
KW Immunogenic region: Human Papilloma Virus; HPV16 E7; diagnosis;  
KW antibodies; vaccines.  
OS Human Papilloma Virus 16 E7.  
PN EP-386734-A.  
PD 12-SEP-1990.  
PF 07-MAR-1990; 104353.  
PR 10-MAR-1989; DE-907721.  
PA (BEHW ) BEHRINGWERKE AG.  
PI Bartsch D, Gissmann L, Muller M;  
DR WPI; 90-276785/37.  
DR P-PSDB; R06684.  
PT New immunogenic regions of Papilloma virus 16 E7 protein - useful  
PT in vaccines and for diagnosis, and new derived antibodies  
PS Example 2; Page 3; 3pp; German.  
CC The sequence extends downstream from nucleotide 667 of a HPV16 DNA  
CC expression bank. HPV16 DNA fragments of ca. 100 bp were blunt-end  
CC ligated into PvuII-cleaved phage vector fd-tet-J6. Recombinant  
CC phages were plated on E. coli K91 and replicated on nitrocellulose  
CC membranes probed with specific sera. 200 recombinants reacted and  
CC 30 of them were sequenced. This sequence was identified from 5  
CC overlapping clones. The sequence encodes immunogenic regions which  
CC are useful in vaccines, to detect specific antibodies against HPV16  
CC E7-protein. Antibodies are also useful diagnostically.  
CC See also Q05881.  
SQ Sequence 17 AA;

Query Match 62.7%; Score 42; DB 2; Length 17;  
Best Local Similarity 40.0%; Pred. NO. 2.85e+02;  
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 4 dgpagaqepd 13

QY 1 DGPTGPEQOE 10

\* \* \*

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.

```

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 21 15:43:59 1999; Maspar time 1.51 Seconds
77.431 Million cell updates/sec
Tabular output not generated.

```

```

Title:
Description:
perfect Score:
Sequence:
>US-09-040-485-8
(1-10) from US09040485.pep
67
1 DGPTGEPQQE 10

```

Scoring table: PAM 150  
Gap 15

Searched: 119857 seqs, 11713122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

```
Database: . a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1
```

Statistics: Mean 14.038; Variance 40.705; scale 0.345

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	47	70.1	1122	2	US-08-619-	Sequence 3, Applicatio	3.20e+01
2	47	70.1	1389	2	US-08-619-	Sequence 5, Applicatio	3.20e+01
3	47	70.1	3567	2	US-07-642-	Sequence 4, Applicatio	3.20e+01
4	44	65.7	243	1	US-07-960-	Sequence 2, Applicatio	7.10e+01
5	44	65.7	243	2	US-08-473-	Sequence 2, Applicatio	7.10e+01
6	44	65.7	243	2	US-08-301-	Sequence 2, Applicatio	7.10e+01
7	44	65.7	243	3	PCT-US93-0	Sequence 2, Applicatio	7.10e+01
8	44	65.7	745	2	US-08-479-	Sequence 2, Applicatio	7.10e+01
9	44	65.7	745	1	US-08-136-	Sequence 2, Applicatio	7.10e+01
10	43	64.2	391	2	US-08-244-	Sequence 10, Applicati	9.24e+01
11	43	64.2	391	2	US-08-389-	Sequence 23, Applicati	9.24e+01
12	43	64.2	391	4	5223423-4	Patent No. 5223423	9.24e+01
13	42	62.7	25	1	US-08-363-	Sequence 2, Applicatio	1.20e+02
14	42	62.7	98	1	US-08-406-	Sequence 6, Applicatio	1.20e+02
15	42	62.7	181	1	US-08-034-	Sequence 12, Applicati	1.20e+02
16	42	62.7	181	1	US-08-034-	Sequence 14, Applicati	1.20e+02
17	42	62.7	253	2	US-08-889-	Sequence 20, Applicati	1.20e+02
18	42	62.7	253	2	US-08-465-	Sequence 20, Applicati	1.20e+02
19	42	62.7	253	2	US-08-459-	Sequence 20, Applicati	1.20e+02
20	42	62.7	263	1	US-08-117-	Sequence 9, Applicatio	1.20e+02
21	42	62.7	1442	2	US-08-316-	Sequence 12, Applicati	1.20e+02
22	42	62.7	1442	3	PCT-US95-0	Sequence 12, Applicati	1.20e+02
23	42	62.7	2237	2	US-08-455-	Sequence 48, Applicati	1.20e+02

## ALIGNMENTS

RESULT 1  
ID US-08-619-198-3 STANDARD; PRT; 1122 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 3, Application US/08619198  
XX  
CC Sequence 3, Application US/08619198  
CC Patent No. 5885831  
CC GENERAL INFORMATION:  
CC APPLICANT: Young, Michael W.  
CC APPLICANT: Sehgal, Amita  
CC APPLICANT: Voshall, Leslie B.  
CC APPLICANT: Price, Jeffrey L.  
CC APPLICANT: Myers, Michael  
CC TITLE OF INVENTION: NUCLEAR LOCALIZATION FACTOR ASSOCIATED  
CC WITH CIRCADIAN RHYTHMS  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Klauber & Jackson  
CC STREET: 411 Hackensack Avenue  
CC CITY: Hackensack  
CC STATE: New Jersey  
CC COUNTRY: USA  
CC ZIP: 07601  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/619,198  
CC FILING DATE: 20-MAR-1996  
CC CLASSIFICATION: 514  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Jackson Esq., David A.  
CC REGISTRATION NUMBER: 26,742  
CC REFERENCE/POCKET NUMBER: 600-1-128A CP1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 201 487-5800  
CC TELEFAX: 201 343-1684  
CC TELEX: 133521  
CC INFORMATION FOR SEQ ID NO: 3:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1122 amino acids



CC APPLICANT: Frisch, Steven M.  
CC TITLE OF INVENTION: Method of Inhibiting Replication of  
CC HYPERPROLIFERATIVE CELLS  
CC NUMBER OF SEQUENCES: 4  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Campbell and Flores  
CC STREET: 4370 La Jolla Village Drive, Suite 700  
CC CITY: San Diego  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 92122  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/960,112B  
CC FILING DATE: 13-OCT-1992  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Campbell, Cathryn A.  
CC REGISTRATION NUMBER: 31,815  
CC REFERENCE/DOCKET NUMBER: P-LJ 9429  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (619) 535-9001  
CC TELEFAX: (619) 535-8949  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 243 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 243 AA; 26451 MW; 304537 CN;  
Query Match 65.7%; Score 44; DB 1; Length 243;  
Best Local Similarity 50.0%; Pred. No. 7.10e+01;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Db 138 EGPVSEPEPE 147  
:|:|:|:|:  
QY 1 DGPTEPQOE 10  
RESULT 5  
ID US-08-473-399B-2 STANDARD; PRT; 243 AA.  
XX  
AC xxxxxx  
DT  
DT  
DE  
DE  
XX  
Sequence 2, Application US/08473399B  
Sequence 2, Application US/08473399B  
Patent No. 5866550  
GENERAL INFORMATION:  
CC APPLICANT: Frisch, Steven M.  
CC TITLE OF INVENTION: REVERSE-TRANSFORMATION OF CANCER CELLS  
CC NUMBER OF SEQUENCES: 4  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Campbell & Flores LLP  
CC STREET: 4370 La Jolla Village Drive, Suite 700  
CC CITY: San Diego  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 92122  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/473,399B  
CC FILING DATE: 06-JUN-1995  
CC CLASSIFICATION: 514  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/960,112  
CC FILING DATE: 13-OCT-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Campbell, Cathryn A.  
CC REGISTRATION NUMBER: 31,815  
CC REFERENCE/DOCKET NUMBER: P-LJ 1697  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (619) 535-9001  
CC TELEFAX: (619) 535-8949  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 243 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 243 AA; 26451 MW; 304537 CN;  
Query Match 65.7%; Score 44; DB 2; Length 243;  
Best Local Similarity 50.0%; Pred. No. 7.10e+01;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Db 138 EGPVSEPEPE 147  
:|:|:|:|:  
QY 1 DGPTEPQOE 10  
RESULT 6  
ID US-08-301-316B-2 STANDARD; PRT; 243 AA.  
XX  
AC xxxxxx  
DT  
DT  
DE  
DE  
XX  
Sequence 2, Application US/08301316B  
Sequence 2, Application US/08301316B  
Patent No. 5776743  
GENERAL INFORMATION:  
CC APPLICANT: Frisch, Steven M.  
CC TITLE OF INVENTION: Method of Sensitizing Tumor Cells with Adenovirus ELA  
CC NUMBER OF SEQUENCES: 4  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Campbell & Flores LLP  
CC STREET: 4370 La Jolla Village Drive, Suite 700  
CC CITY: San Diego  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 92122  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/301,316B  
CC FILING DATE: 06-SEP-1994  
CC CLASSIFICATION: 424  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Campbell, Cathryn A.  
CC REGISTRATION NUMBER: 31,815  
CC REFERENCE/DOCKET NUMBER: P-LJ 1115  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (619) 535-9001  
CC TELEFAX: (619) 535-8949  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 243 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear

```
CC MOLECULE TYPE: protein
SQ SEQUENCE 243 AA; 26451 MW; 304537 CN;

Query Match 65.7%; Score 44; DB 2; Length 243;
Best Local Similarity 50.0%; Pred. No. 7.10e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 138 EGPVSEPEPE 147
QY 1 DGPTGEPOQE 10

RESULT 7
ID PCT-US93-09774-2 STANDARD; PRT; 243 AA.
XX
AC xxxxxx
XX
DT
DE
DE Sequence 2, Application PC/TUS9309774
XX
CC Sequence 2, Application PC/TUS9309774
CC GENERAL INFORMATION:
CC APPLICANT: LA JOLLA CANCER RESEARCH FOUNDATION
CC TITLE OF INVENTION: METHOD OF INHIBITING REPLICATION OF
CC TITLE OF INVENTION: HYPERPROLIFERATIVE CELLS
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: CAMPBELL AND FLORES
CC STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CC CITY: SAN DIEGO
CC STATE: CALIFORNIA
CC COUNTRY: UNITED STATES
CC ZIP: 92122
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/09774
CC FILING DATE: 12-OCT-1992
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: PERKINS, SUSAN M.
CC REGISTRATION NUMBER: 36,405
CC REFERENCE/DOCKET NUMBER: FP-LJ 9770
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 619-535-9001
CC TELEFAX: 619-535-8949
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 243 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 243 AA; 26451 MW; 304537 CN;

Query Match 65.7%; Score 44; DB 3; Length 243;
Best Local Similarity 50.0%; Pred. No. 7.10e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 138 EGPVSEPEPE 147
QY 1 DGPTGEPOQE 10

RESULT 8
ID US-08-479-403-2 STANDARD; PRT; 745 AA.
XX
AC xxxxxx
XX
DT
DE
DE Sequence 2, Application US/08136277
XX
CC Sequence 2, Application US/08136277
CC Patent No. 5644045
CC GENERAL INFORMATION:
CC APPLICANT: MANDEL, Jean-Louis
CC APPLICANT: AUBOURG, Patrick
CC APPLICANT: MOSSER, Jean
CC APPLICANT: SARDE, Claude
CC TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
CC NUMBER OF SEQUENCES: 23

XX Sequence 2, Application US/08479403
DE
XX Sequence 2, Application US/08479403
CC Patent No. 5869039
CC GENERAL INFORMATION:
CC APPLICANT: MANDEL, Jean-Louis
CC APPLICANT: AUBOURG, Patrick
CC APPLICANT: MOSSER, Jean
CC APPLICANT: SARDE, Claude
CC TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
CC NUMBER OF SEQUENCES: 23
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Young & Thompson
CC STREET: 745 South 23rd Street
CC CITY: Arlington
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22202
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/479,403
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:
CC NAME: PATCH, Andrew J.
CC REGISTRATION NUMBER: 32,925
CC REFERENCE/DOCKET NUMBER: B2272DIV
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-521-2297
CC TELEFAX: 703-685-0573
CC TELEX: 248425 EMBON
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 745 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 745 AA; 82908 MW; 2775045 CN;

Query Match 65.7%; Score 44; DB 2; Length 745;
Best Local Similarity 60.0%; Pred. No. 7.10e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPTQE 56
QY 1 DGPTGEPOQE 10

RESULT 9
ID US-08-136-277-2 STANDARD; PRT; 745 AA.
XX
AC xxxxxx
XX
DT
DE
DE Sequence 2, Application US/08136277
XX
CC Sequence 2, Application US/08136277
CC Patent No. 5644045
CC GENERAL INFORMATION:
CC APPLICANT: MANDEL, Jean-Louis
CC APPLICANT: AUBOURG, Patrick
CC APPLICANT: MOSSER, Jean
CC APPLICANT: SARDE, Claude
CC TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
CC NUMBER OF SEQUENCES: 23
```

CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Young & Thompson  
CC STREET: 745 South 23rd Street  
CC CITY: Arlington  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22202  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/136,277  
CC FILING DATE: 15-OCT-1993  
CC CLASSIFICATION: 424  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: PATCH, Andrew J.  
CC REGISTRATION NUMBER: 32,925  
CC REFERENCE/DOCKET NUMBER: B2272  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 703-521-2297  
CC TELEFAX: 703-685-0573  
CC TELEX: 248425 EMBON  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 745 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 745 AA; 82908 MW; 2775045 CN;  
SQ  
  
Query Match 65.7%; Score 44; DB 1; Length 745;  
Best Local Similarity 60.0%; Pred. No. 7.10e+01;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Db 47 QAPAGEPTQE 56  
QY 1 DGTPGEPTQE 10  
  
RESULT 10  
ID US-08-244-951A-10 STANDARD; PRT; 391 AA.  
XX AC xxxxxx  
XX DT  
XX DE  
XX SEQUENCE 10, Application US/08244951A  
XX  
Sequence 10, Application US/08244951A  
Patent No. 5843779  
GENERAL INFORMATION:  
CC APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;  
CC APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;  
CC TITLE OF INVENTION: MONOCLONAL ANTIBODIES  
CC TITLE OF INVENTION: DIRECTED AGAINST THE MICROTUBULE-ASSOCIATED  
CC TITLE OF INVENTION: PROTEIN TAU, HYBRIDOMAS SECRETING THESE  
CC TITLE OF INVENTION: ANTIBODIES, ANTIGEN RECOGNITION BY THESE  
CC TITLE OF INVENTION: MONOCLONAL ANTIBODIES AND THEIR APPLICATIONS  
CC NUMBER OF SEQUENCES: 10  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: BIERMAN & MUSERLIAN  
CC STREET: 600 THIRD AVENUE  
CC CITY: NEW YORK  
CC STATE: NEW YORK  
CC COUNTRY: USA  
CC ZIP: 10016  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: FLOPPY DISK  
CC COMPUTER: IBM PC COMPATIBLE  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: ASCII  
CC  
CC  
CC

CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/244,951A  
CC FILING DATE: 19-JAN-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/EP93/03499  
CC FILING DATE: 10-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: EP/92/403403.6  
CC FILING DATE: 14-DEC-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: CHARLES A. MUSERLIAN  
CC REGISTRATION NUMBER: 19,683  
CC REFERENCE/DOCKET NUMBER: 410.003A  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 661-8000  
CC TELEFAX: (212) 661-8002  
CC INFORMATION FOR SEQ ID NO: 10:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391  
CC TYPE: Amino Acid  
CC STRANDEDNESS: Unknown  
CC TOPOLOGY: Unknown  
CC FEATURE:  
CC NAME/KEY: mTHEMPH-tau fusion protein  
CC SEQUENCE 391 AA; 41065 MW; 834762 CN;  
SQ  
  
Query Match 64.2%; Score 43; DB 2; Length 391;  
Best Local Similarity 66.7%; Pred. No. 9.24e+01;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Db 38 GPMAEPQOE 46  
QY 2 GPTGEPTQE 10  
  
RESULT 11  
ID US-08-389-011-23 STANDARD; PRT; 391 AA.  
XX AC xxxxxx  
XX DT  
XX DE  
XX SEQUENCE 23, Application US/08389011  
XX  
Sequence 23, Application US/08389011  
Patent No. 5861257  
GENERAL INFORMATION:  
CC APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;  
CC APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;  
CC TITLE OF INVENTION: MONOCLONAL ANTIBODIES  
CC TITLE OF INVENTION: DIRECTED AGAINST THE MICROTUBULE-ASSOCIATED  
CC TITLE OF INVENTION: PROTEIN TAU, HYBRIDOMAS SECRETING THESE  
CC TITLE OF INVENTION: ANTIBODIES, ANTIGEN RECOGNITION BY THESE  
CC TITLE OF INVENTION: MONOCLONAL ANTIBODIES AND THEIR APPLICATIONS  
CC NUMBER OF SEQUENCES: 24  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: BIERMAN & MUSERLIAN  
CC STREET: 600 THIRD AVENUE  
CC CITY: NEW YORK  
CC STATE: NEW YORK  
CC COUNTRY: USA  
CC ZIP: 10016  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: FLOPPY DISK  
CC COMPUTER: IBM PC COMPATIBLE  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: ASCII  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/389,011  
CC FILING DATE: 13-FEB-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC

CC APPLICATION NUMBER: 08/403,917  
CC FILING DATE: 19-JAN-1995  
CC PRIOR APPLICATION NUMBER: 08/403,916  
CC FILING DATE: 19-JAN-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/244,951  
CC FILING DATE: 13-JUN-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/EP93/03499  
CC FILING DATE: 10-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: EP/92/403403.6  
CC FILING DATE: 14-DEC-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: CHARLES A. MUSERLIAN  
CC REGISTRATION NUMBER: 19,683  
CC REFERENCE/DOCKET NUMBER: 410.003-1-CON  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 661-8000  
CC TELEFAX: (212) 661-8002  
CC INFORMATION FOR SEQ ID NO: 23:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 391  
CC TYPE: Amino Acid  
CC STRANDEDNESS: Unknown  
CC TOPOLOGY: Unknown  
CC SEQUENCE 391 AA; 41065 MW; 834762 CN;  
  
Query Match 64.2%; Score 43; DB 2; Length 391;  
Best Local Similarity 66.7%; Pred. No. 9.24e+01;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Db 38 GPMAEPQOE 46  
QY 2 GPTGEPPQOE 10  
  
RESULT 12  
ID 5223423-4 STANDARD; PRT: 3336 AA.  
XX  
AC xxxxxx  
DT 01-JAN-1900  
XX Patent No. 5223423.  
XX Patent No. 5223423  
XX APPLICANT: FRANCHINI, GENOVEFFA, WONG-STAAAL, FLOSSIE;  
XX GALLO, ROBERT  
CC TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT  
CC HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY  
CC NUMBER OF SEQUENCES: 4  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/331,212  
CC FILING DATE: 03-31-1989  
CC SEQ ID NO: 4:  
CC LENGTH: 3080  
CC SEQUENCE 3336 AA; 377678 MW; 61548087 CN;  
  
Query Match 64.2%; Score 43; DB 4; Length 3080;  
Best Local Similarity 75.0%; Pred. No. 9.24e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Db 2799 PHGEPPQOE 2806  
QY 3 PTGEPPQOE 10  
  
RESULT 13  
ID US-08-363-586-2 STANDARD; PRT: 25 AA.  
XX  
AC xxxxxx

XX  
DT  
XX  
DE  
XX  
CC Sequence 2, Application US/08363586  
CC Sequence 2, Application US/08363586  
CC Patent No. 5629161  
CC GENERAL INFORMATION:  
CC APPLICANT: Mueller, Martin  
CC APPLICANT: Gissmann, Lutz  
CC TITLE OF INVENTION: Use of HPV-16 E6 and E7-Gene Derived  
CC TITLE OF INVENTION: Peptides for the Diagnostic Purpose  
CC NUMBER OF SEQUENCES: 4  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
CC ADDRESSEE: Dunner  
CC STREET: 1300 I Street, N.W.  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005-3315  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/363,586  
CC FILING DATE: 23-DEC-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/909,296  
CC FILING DATE: 09-JUL-1992  
CC APPLICATION NUMBER: EP 91111720.8  
CC FILING DATE: 13-JUL-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Wadler, Linda A.  
CC REGISTRATION NUMBER: 33,218  
CC REFERENCE/DOCKET NUMBER: 02481-1195-00000  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202-408-4000  
CC TELEFAX: 202-408-4400  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 25 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC SEQUENCE 25 AA; 2746 MW; 2319 CN;  
  
Query Match 62.7%; Score 42; DB 1; Length 25;  
Best Local Similarity 40.0%; Pred. No. 1.20e+02;  
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
  
Db 11 DGPAGQAEPD 20  
QY 1 DGPTGEPPQOE 10  
  
RESULT 14  
ID US-08-406-248-6 STANDARD; PRT: 98 AA.  
XX  
AC xxxxxx  
DT  
XX  
DE  
XX  
CC Sequence 6, Application US/08406248  
CC Sequence 6, Application US/08406248  
CC Patent No. 5736318  
CC GENERAL INFORMATION:  
CC APPLICANT: Munger, Karl  
CC APPLICANT: Jones, D. Leanne



CC TITLE OF INVENTION: METHOD AND KIT FOR EVALUATING  
CC TITLE OF INVENTION: TRANSFORMED CELLS  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Ann-Louise Kerner, Ph.D., Lappin & Kusmer  
CC STREET: 200 State Street  
CC CITY: Boston  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02109  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/406,248

CC FILING DATE:

CC CLASSIFICATION: 436

CC ATTORNEY/AGENT INFORMATION:

CC NAME: McDaniels, Patricia A.

CC REGISTRATION NUMBER: 33,194

CC REFERENCE/DOCKET NUMBER: HAZ-011

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 617-330-1300

CC TELEFAX: 617-330-1311

CC INFORMATION FOR SEQ ID NO: 6:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 98 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 98 AA; 11022 MW; 49177 CN;

Query Match 62.7%; Score 42; DB 1; Length 98;  
Best Local Similarity 40.0%; Pred. No. 1.20e+02;  
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 39 DGPAGQAPD 48

QY 1 DGPTGPQQE 10

## RESULT 15

ID US-08-034-245-12 STANDARD; PRT; 181 AA.

XX AC xxxxxx

XX AC xxxxxx

XX DT

XX DE

XX Sequence 12, Application US/08034245

XX Sequence 12, Application US/08034245

CC Patent No. 5504197

CC GENERAL INFORMATION:

CC APPLICANT: Schubert, David

CC APPLICANT: Fisher, Wolfgang H.

CC TITLE OF INVENTION: NEUTROTROPHIC GROWTH FACTOR AND METHODS

CC TITLE OF INVENTION: OF TREATMENT

CC NUMBER OF SEQUENCES: 14

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark

CC STREET: 444 South Flower Street, Suite 2000

CC CITY: Los Angeles

CC STATE: CA

CC COUNTRY: USA

CC ZIP: 90071

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/034,245  
CC FILING DATE: 19930322  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/590,359  
CC FILING DATE: 27-SEP-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/541,276  
CC FILING DATE: 20-JUN-1990  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Reiter, Stephen E.  
CC REGISTRATION NUMBER: 31,192  
CC REFERENCE/DOCKET NUMBER: P41 9369  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 619-546-4737  
CC TELEFAX: 619-546-9392  
CC INFORMATION FOR SEQ ID NO: 12:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 181 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 181 AA; 20536 MW; 137417 CN;

Query Match 62.7%; Score 42; DB 1; Length 181;  
Best Local Similarity 50.0%; Pred. No. 1.20e+02;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 45 DGASGDPKKE 54

QY 1 DGPTGPQQE 10

Search completed: Thu Oct 21 15:44:06 1999  
Job time : 7 secs.

\*\*\*\*\*  
M P E R L  
\*\*\*\*\* (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 21 15:41:52 1999; MasPar time 3.22 Seconds  
Tabular output not generated. 124.586 Million cell updates/sec

Title: >US-09-040-485-8  
Description: (1-10) from US09040485.p  
Perfect Score: 67  
Sequence: 1 DGPTGEPQOE 10

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 20.467; Variance 25.896; scale 0.790

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	67	100.0	aspartyl beta-hydroxy	1.17e-04
2	50	74.6	25.5K membrane protei	1.08e+00
3	48	71.6	hypothetical protein	2.89e+00
4	47	70.1	tim (timeless) protei	4.69e+00
5	47	70.1	erythronolide synthas	4.69e+00
6	45	67.2	collagen alpha 1(I) c	1.21e+01
7	45	67.2	probable membrane pro	1.21e+01
8	45	67.2	MHC class II histocom	1.21e+01
9	45	67.2	p53-binding protein 1	1.21e+01
10	45	67.2	collagen alpha 1(II)	1.21e+01
11	45	67.2	neurogenic locus mam	1.21e+01
12	44	65.7	kinesin-related prote	1.93e+01
13	44	65.7	probable transport pr	1.93e+01
14	44	65.7	adrenoleukodystrophy	1.93e+01
15	44	65.7	1-phosphatidylinosito	1.93e+01
16	44	65.7	chemotaxis protein ch	1.93e+01
17	44	65.7	KIAA0647 protein - hu	1.93e+01
18	44	65.7	NF-180 - sea lamprey	1.93e+01
19	44	65.7	B120 protein - human	1.93e+01
20	44	65.7	hypothetical protein	1.93e+01
21	44	65.7	collagen alpha 1(XV)	1.93e+01
22	43	64.2	major histocompatibil	3.04e+01
23	43	64.2	major histocompatibil	3.04e+01

24	43	64.2	264	2	B30582	MHC class II histocom	3.04e+01
25	43	64.2	449	2	A41520	chromogranin A precu	3.04e+01
26	43	64.2	944	2	T00265	respiratory burst oxi	3.04e+01
27	43	64.2	1028	1	CGHUA	collagen alpha 1(VI)	3.04e+01
28	43	64.2	1670	1	CGHUB	collagen alpha 3(IV)	3.04e+01
29	43	64.2	1707	2	A33526	collagen alpha 2(IV)	3.04e+01
30	43	64.2	1841	1	JDMU1	DNA-directed RNA poly	3.04e+01
31	43	64.2	1860	1	JDMU2	DNA-directed RNA poly	3.04e+01
32	42	62.7	181	2	S62782	heat- and acid-stable	4.77e+01
33	42	62.7	235	2	S3249	transcription factor	4.77e+01
34	42	62.7	286	2	S34665	collagen, cuticular	4.77e+01
35	42	62.7	323	2	A61396	collagen alpha 1(II)	4.77e+01
36	42	62.7	324	2	S58061	amelin 2 - rat	4.77e+01
37	42	62.7	384	2	T51267	collagenase (EC 3.4.2	4.77e+01
38	42	62.7	407	2	S58062	amelin 1 - rat	4.77e+01
39	42	62.7	496	2	S26402	homeotic protein Hox	4.77e+01
40	42	62.7	674	2	S13301	collagen alpha 1(X) c	4.77e+01
41	42	62.7	736	2	S47044	MALDP protein - mouse	4.77e+01
42	42	62.7	1019	2	A32856	collagen alpha 1(VI)	4.77e+01
43	42	62.7	1365	2	A41483	glucosyltransferase (	4.77e+01
44	42	62.7	1487	2	B41182	collagen alpha 1(II)	4.77e+01
45	42	62.7	2541	2	S11661	talin - mouse	4.77e+01

ALIGNMENTS

RESULT 1  
ENTRY I38423 #type complete  
TITLE aspartyl beta-hydroxylase - human  
ORGANISM #formal name Homo sapiens #common name man  
DATE 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 10-Jul-1998

ACCESSIONS I38423  
REFERENCE I38423  
#authors Koriath, F.; Gieffers, C.; Frey, J.  
#journal Gene (1994) 150:395-399  
#title Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase.  
#cross-references MIM:95121937  
#accession I38423

##status preliminary; translated from GB/EMBL/DBDJ  
##molecule\_type mRNA  
##residues 1-757 #label RES  
##cross-references EMBL:003109; NID:G458031; PID:G458032  
CLASSIFICATION #superfamily peptidase-aspartate beta-dioxygenase; tetratricopeptide repeat homology

FEATURE 54-75 #domain transmembrane #status predicted #label TRM  
SUMMARY #length 757 #molecular-weight 85498 #checksum 2143

Query Match 100.0%; Score 67; DB 2; Length 757;  
Best Local Similarity 100.0%; Pred. No. 1.17e-04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 173 DGPTGEPQOE 182  
QY 1 DGPTGEPQOE 10

RESULT 2  
ENTRY MMBE13 #type complete  
TITLE 25.5K membrane protein - ictaluriid herpesvirus 1 (strain auburn 1)  
ORGANISM #formal name ictaluriid herpesvirus 1  
#note host ictalurus punctatus (channel catfish)  
DATE 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 05-Sep-1997

ACCESSIONS I36786  
REFERENCE I36786  
#authors Davison, A.J.  
#submission submitted to GenBank, January 1992  
#description Channel catfish virus: a new type of herpesvirus.  
#accession I36786

```

##molecule_type DNA
##residues 1-232 ##label DAV
##cross-references GB:M75136; NID:g331209; PID:g331218
REFERENCE A39447
#authors Davison, A.J.
#journal Virology (1992) 186:9-14
#title Channel catfish virus: a new type of herpesvirus.
#cross-references MUID:92087490
#contents annotation
#note neither amino acid nor nucleotide sequence is given
GENETICS
#gene 8
#status preliminary
#superfamily ictaluriid herpesvirus 25.5K membrane protein
KEYWORDS transmembrane protein
FEATURE
#domain transmembrane #status predicted #label TM1\
#domain transmembrane #status predicted #label TM2
SUMMARY #length 232 #molecular-weight 25466 #checksum 8650
Query Match 74.6%; Score 50; DB 1; Length 232;
Best Local Similarity 50.0%; Pred. No. 1.08e+00;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 88 ESPTGEPHRD 97
QY 1 DGPTGEPQOE 10
RESULT 3
ENTRY S75052 #type complete
TITLE hypothetical protein sll1601 - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
ACCESSIONS S75052
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hiroseawa, M.; Sugiyura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#cross-references MUID:97061201
#accession S75052
#status preliminary
##molecule_type DNA
##residues 1-414 ##label KAN
##cross-references EMBL:D50910; GB:AB001339; NID:g1652956; PID:d1018647;
#note the nucleotide sequence was submitted to the EMBL Data Library, June 1996
SUMMARY #length 414 #molecular-weight 44810 #checksum 1664
Query Match 71.6%; Score 48; DB 2; Length 414;
Best Local Similarity 60.0%; Pred. No. 2.89e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 318 EGPTAAQOO 327
QY 1 DGPTGEPQOE 10
RESULT 4
ENTRY A57655 #type complete
TITLE tim (timeless) protein - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster

```

```

DATE 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 17-Mar-1999
ACCESSIONS A57655
REFERENCE A57655
#authors Myers, M.P.; Wager-Smith, K.; Wesley, C.S.; Young, M.W.; Sehgal, A.
#journal Science (1995) 270:805-808
#title Positional cloning and sequence analysis of the Drosophila clock gene, timeless.
#cross-references MUID:96055118
#accession A57655
#status preliminary; nucleic acid sequence not shown
##molecule_type DNA
##residues 1-1388 ##label MYE
##cross-references GB:U37018
GENETICS
#gene tim
#accession A57655
SUMMARY ##cross-references FlyBase:FBgn0014396
#length 1388 #molecular-weight 155537 #checksum 3459
Query Match 70.1%; Score 47; DB 2; Length 1388;
Best Local Similarity 60.0%; Pred. No. 4.69e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 510 DGPOGKPOHQ 519
QY 1 DGPTGEPQOE 10
RESULT 5
ENTRY S23070 #type complete
TITLE erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea
ORGANISM 6-deoxyerythronolide B synthase II
ALTERNATE_NAMES #formal_name Saccharopolyspora erythraea
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 24-Mar-1999
ACCESSIONS S23070; S22011; S23205
REFERENCE S23070
#authors Bevitte, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
#journal Eur. J. Biochem. (1992) 204:39-49
#title 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme.
#accession S23070
##molecule_type DNA
##residues 1-3573 ##label BEV1
##cross-references EMBL:X62569
##experimental_source strain NRRL 2338
REFERENCE S22011
#authors Bevitte, D.J.
#submission submitted to the EMBL Data Library, September 1991
#accession S22011
##molecule_type DNA
##residues 1-184, 'I', 186-301, 'S', 303-521, 523-658, 'A', 660-993, 1001-1212, 'H', 1214-1392, 1394-2481, 'V', 2482-2827, 'P', 2829-2833, 'L', 2835-2856, 2858-2907, 'A', 2908-3135, 'K', 3137-3166, 'H', 3168-3176, 'L', 3177-3479, 'DH', 3480-3572 ##label BEV2
##cross-references EMBL:X62569; NID:g46977; PID:g581651
REFERENCE S23103
#authors Caffrey, P.; Bevitte, D.J.; Staunton, J.; Leadlay, P.F.
#journal FEBS Lett. (1992) 304:225-228
#title Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from Saccharopolyspora erythraea.
#cross-references MUID:92316235
#accession S23205
##molecule_type protein
##residues 2-12, 'XXX' ##label CAF
##experimental_source strain CA340
GENETICS

```

```
#gene      eryA
#start_codon  GTG
#description catalyzes the construction of a polyketide chain, which is
              then cyclised to form 14-membered lacton ring of
              6-deoxyerythronolide B
#pathway    erythromycin biosynthesis
CLASSIFICATION
#superfamily [acyl-carrier-protein] S-malonyltransferase
homology; 3-oxoacyl-[acyl-carrier-protein] synthase I
homology; acyl carrier protein homology; long-chain alcohol
dehydrogenase homology; short-chain alcohol dehydrogenase
homology
KEYWORDS    acyltransferase; antibiotic biosynthesis; multifunctional
              enzyme
FEATURE
52-453      #domain 3-oxoacyl-[acyl-carrier-protein] synthase I
              homology #label OAS1\
561-843      #domain [acyl-carrier-protein] S-malonyltransferase
              homology #label AMT1\
1140-1308    #domain short-chain alcohol dehydrogenase homology
              #label SAD2\
1404-1475    #domain acyl carrier protein homology #label ACP1\
1519-1919    #domain 3-oxoacyl-[acyl-carrier-protein] synthase I
              homology #label OAS2\
2023-2305    #domain [acyl-carrier-protein] S-malonyltransferase
              homology #label AMT2\
2857-3131    #domain long-chain alcohol dehydrogenase homology #label
              LADH\
3149-3327    #domain short-chain alcohol dehydrogenase homology
              #label SADH\
3420-3493    #domain acyl carrier protein homology #label ACP2
              #length 3573 #molecular-weight 375258 #checksum 9593
SUMMARY
Query Match      70.18; Score 47; DB 2; Length 3573;
Best Local Similarity 87.5%; Pred. No. 4.69e+00;
Matches          7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1384 GPTGPEAQ 1391
QY 2 GPTGPEQQ 9

RESULT 6
ENTRY A05249 #type fragment
TITLE collagen alpha 1(I) chain precursor - baboon (fragment)
ORGANISM #formal_name Papio sp. #common_name baboon
DATE 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change
21-Jul-1995
ACCESSIONS A05249
REFERENCE A02078
#authors Epstein Jr., E.H.; Scott, R.D.; Miller, E.J.; Piez, K.A.
#journal J. Biol. Chem. (1971) 246:1718-1724
#title Isolation and characterization of the peptides derived from
soluble human and baboon skin collagen after cyanogen
bromide cleavage.
#cross-references MUID:71134791
#contents CNBr0-1, CNBr2, CNBr4, CNBr5, composition
#accession A05249
#molecule_type protein
#residues 1-140 #label EPS
#experimental_source skin
COMMENT Prolines at the third position of the tripeptide repeating unit
(G-X-Y) are hydroxylated in some or all of the chains.
CLASSIFICATION
#superfamily collagen alpha 1(I) chain; fibrillar collagen
carboxyl-terminal homology; von Willebrand factor type C
repeat homology
KEYWORDS coiled coil; extracellular matrix; glycoprotein; skin;
trimer; triple helix
SUMMARY #length 140 #checksum 7697

Query Match      67.28; Score 45; DB 2; Length 140;
Best Local Similarity 44.48; Pred. No. 1.21e+01;
Matches          4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

```
Db 41 ZGPPGZPCZ 49
QY 1 DGPTGPEQQ 9

RESULT 7
ENTRY G64831 #type complete
TITLE probable membrane protein ybcC - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
13-Sep-1998
ACCESSIONS G64831
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession G64831
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-259 #label BLAT
#cross-references GB:AE000194; GB:U00096; NID:g1787148; PID:g1787150;
WGSP:b0920
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene ybcC
KEYWORDS transmembrane protein
FEATURE 13-29
39-55
SUMMARY #length 259 #molecular-weight 28666 #checksum 3266

Query Match      67.28; Score 45; DB 2; Length 259;
Best Local Similarity 66.7%; Pred. No. 1.21e+01;
Matches          6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 251 GSGGEPQEQ 259
QY 2 GPTGPEQQE 10

RESULT 8
ENTRY A45838 #type complete
TITLE MHC class II histocompatibility antigen B-LBII - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 03-Jun-1993 #sequence_revision 30-Sep-1993 #text_change
08-Sep-1997
ACCESSIONS A45838
REFERENCE A45838
#authors Zoorob, R.; Behar, G.; Kroemer, G.; Auffray, C.
#journal Immunogenetics (1990) 31:179-187
#title Organization of a functional chicken class II B gene.
#cross-references MUID:90202026
#accession A45838
#status preliminary
#molecule_type DNA
#residues 1-263 #label ZOO
#cross-references GB:M29763; NID:g212297; PID:g212298
#note the authors translated the codon GTG for residue 112 as
Asp
CLASSIFICATION #superfamily class II histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 263 #molecular-weight 29112 #checksum 5224

Query Match      67.28; Score 45; DB 2; Length 263;
Best Local Similarity 70.08; Pred. No. 1.21e+01;
Matches          7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```

Db 76 DPLGEPOAE 85
    1 1 1 1 1 1 1 1
Qy 1 DPLGEPOAE 10

RESULT 9
ENTRY #type fragment
TITLE p53-binding protein 1 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change
ACCESSIONS I38604
REFERENCE I38604
#authors Iwabuchi, K.; Bartel, P.L.; Li, B.; Mairaccino, R.; Fields, S.
#journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:6098-6102
#title Two cellular proteins that bind to wild-type but not mutant p53.
#cross-references MUID:94286584
#accession I38604
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-1027 #label RES
#cross-references EMBL:U09477; NID:g488591; PID:g488592
SUMMARY #length 1027 #checksum 8683

Query Match 67.2%; Score 45; DB 2; Length 1027;
Best Local Similarity 60.0%; Pred. No. 1.21e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 808 DPTGSSEE 817
    1 1 1 1 1 1 1 1
Qy 1 DPTGSSEE 10

RESULT 10
ENTRY #type complete
TITLE collagen alpha 1(II) chain precursor - human
ALTERNATE_NAMES procollagen alpha 1(II) chain
CONTAINS chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen alpha 1(II) chain precursor splice form 2; collagen alpha 3(XI) chain
ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-May-1986 #sequence_revision 01-Sep-1995 #text_change
ACCESSIONS A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S63514; S04892; S05000; A44309; S16502; A02858; A27280; A57033; A21733; B21733; A24561; S59491; I84453; I37250; I37251; I37252; I37253; I37254; I55338; I59335; I61910
REFERENCE A38513
#authors Ryan, M.C.; Sieraski, M.; Sandell, L.J.
#journal Genomics (1990) 8:41-48
#title The human type II procollagen gene: identification of an additional protein-coding domain and location of potential regulatory sequences in the promoter and first intron.
#cross-references MUID:91184811
#accession A38513
#molecule_type DNA
#residues 1-103 #label RYA
#cross-references GB:M60299; NID:gl80883; PID:gl8088
REFERENCE S06715
#authors Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
#journal Nucleic Acids Res. (1989) 17:9473
#title Nucleotide sequence of the full length cDNA encoding for human type II procollagen.
#cross-references MUID:90067946
#accession S06715
#molecule_type mRNA
#residues 1-28; 'R', 99-1487 #label S02
#cross-references EMBL:X16468; NID:g29515; PID:g29516
#note alternative splice form 1
REFERENCE S24270

```

---

```

#authors Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
#journal Biochem. J. (1992) 285:287-294
#title Structural analysis of the regulatory elements of the type-II procollagen gene. Conservation of promoter and first intron sequences between human and mouse.
#cross-references MUID:92344585
#accession S24270
#status translation not shown
#molecule_type DNA
#residues 1-28 #label VIK
#cross-references EMBL:X58709; GB:S40537; NID:g35659
REFERENCE A24828
#authors Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
#journal Gene (1986) 44:111-116
#title Promoter region of the human pro-alpha-1-(II)-collagen gene.
#cross-references MUID:87031574
#accession A24828
#molecule_type DNA
#residues 1-8, 'T', 10-28 #label NUN
#cross-references GB:M25698; NID:gl80872; PID:g553237
REFERENCE S06496
#authors Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
#journal Biochem. J. (1989) 262:521-528
#title Structure of cDNA clones coding for human type II procollagen. The alpha-1(II) chain is more similar to the alpha-1(I) chain than two other alpha chains of fibrillar collagens.
#cross-references MUID:90026318
#accession S06496
#molecule_type mRNA
#residues 7-28; 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1229 #label BAL
#cross-references EMBL:X16711; NID:g30040; PID:g30041
#note alternative splice form 1
REFERENCE A35428
#authors Ryan, M.C.; Sandell, L.J.
#journal J. Biol. Chem. (1990) 265:10334-10339
#title Differential expression of a cysteine-rich domain in the amino-terminal propeptide of type II (cartilage) procollagen by alternative splicing of mRNA.
#cross-references MUID:90285153
#accession A35428
#status not compared with conceptual translation
#molecule_type mRNA
#residues 27-81, 'L', 83-103 #label RYA2
#note alternative splice form 2; splicing appears to be under developmental regulation
REFERENCE A30147
#authors Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
#journal Genomics (1989) 4:438-441
#title Organization of the exons coding for pro alpha-1(II) collagen N-propeptide confirms a distinct evolutionary history of this domain of the fibrillar collagen genes.
#cross-references MUID:89233138
#accession A30147
#molecule_type DNA
#residues 104-157, 'P', 159-236 #label SUM
#cross-references GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GB:M64345; NID:gl80867; PID:gl80869
REFERENCE A94227
#authors Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:6565-6568
#title Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis associated with a mild chondrodysplasia.
#cross-references MUID:90370826
#accession A33116
#molecule_type DNA
#residues 171-172, 'C', 174-175 #label ALA

```

```

#note      mutant sequence from a family with family with primary
#REFERENCE  generalized osteoarthritis and mild chondrodysplasia
#authors   S63514
#journal   Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene
#title     de la Porte, P.; Herbage, D.
#cross-references EMBL:13783; NID:g30037; PID:g930050
#accession S05000
#molecule_type DNA
#residues  630-640,'A',642-785 ##label VIK2
#cross-references EMBL:X16158; NID:g29951
#REFERENCE  A44309
#authors   Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin,
#journal   D.L.; Cohn, D.H.; Eyre, D.R.
#title     J. Biol. Chem. (1992) 267:22522-22526
#cross-references EMBL:117273; NID:g180812
#accession A44309
#status    nucleic acid sequence not shown; not compared with
#molecule_type DNA; mRNA
#residues  752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',
#cross-references GB:L00977; NID:g180812
#note      sequence extracted from NCBI backbone (NCBI:117273);
#note      parts of this sequence were determined by protein
#note      sequencing
#note      mutant sequence associated with perinatal lethal
#note      hypochondrogenesis
#REFERENCE  S16502
#authors   Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1990) 87:3889-3893
#title     Tandem duplication within a type II collagen gene (COL2A1)
#cross-references EMBL:M37126; NID:g180808; PID:g180809
#accession S16502
#molecule_type DNA
#residues  1164-1184,'GPSGKGANGIPGP',1185-1199 ##label TIL
#cross-references EMBL:M37126; NID:g180808; PID:g180809
#note      mutant sequence from a patient with spondyloepiphyseal
#note      dysplasia
#REFERENCE  A02858
#authors   Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.;
#journal   Proc. Natl. Acad. Sci. U.S.A. (1985) 82:2555-2559
#title     Identification and characterization of the human type II
#cross-references EMBL:85190534
#accession A02858
#molecule_type DNA
#residues  1032-1056,'N',1058-1068,'T',1070-1487 ##label CHE
#cross-references GB:J00116; NID:g180395; PID:g180396
#REFERENCE  A27280

```

```

#authors   Elima, K.; Vuorio, T.; Vuorio, E.
#journal   Nucleic Acids Res. (1987) 15:9499-9504
#title     Determination of the single polyadenylation site of the human
#cross-references MUID:88067771
#accession A27280
#molecule_type DNA; mRNA
#residues  1175-1487 ##label ELI
#cross-references EMBL:X06268; NID:g30096; PID:g30097
#experimental_source fetal epiphyseal cartilage
#REFERENCE  A57033
#authors   van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
#journal   Biochem. J. (1986) 237:923-925
#title     Chondrocalcin is identical with the C-propeptide of type II
#accession A57033
#molecule_type protein
...
Note: remainder of annotations omitted.
Query Match      67.2%; Score 45; DB 1; Length 1487;
Best Local Similarity 66.7%; Pred. No. 1.21e+01;
Matches          6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db              999 GPSGEPGQ 1007
QY              11:11111
QY              2 GPTGEPQOE 10
RESULT 11
ENTRY   A33106 #type complete
TITLE   neurogenic locus mam protein - fruit fly (Drosophila
        melanogaster)
ALTERNATE_NAMES mastermind protein
ORGANISM #formal_name Drosophila melanogaster
DATE      07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
        24-Sep-1998
ACCESSIONS A36391; A33106; S13514
REFERENCE  A36391
#authors   Smoller, D.; Friedel, C.; Schmid, A.; Bettler, D.; Lam, L.;
#journal   Yedvobnick, B.
#title     Genes Dev. (1990) 4:1688-1700
#cross-references MUID:91065516
#accession A36391
#status    preliminary
#molecule_type mRNA
#residues  #cross-references GB:X54251; NID:g8203; PID:g8204
#note      strain Canton S
GENETICS  FlyBase:mam
#gene      #cross-references FlyBase:FBgn0002643
SUMMARY   #length 1596 #molecular_weight 167717 #checksum 4406
Query Match      67.2%; Score 45; DB 2; Length 1596;
Best Local Similarity 66.7%; Pred. No. 1.21e+01;
Matches          6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db              1317 GPMGPPQOQ 1325
QY              11:11111
QY              2 GPTGEPQOE 10
RESULT 12
ENTRY   I54523 #type fragment
TITLE   kinesin-related protein - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE      07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change
        05-Jun-1998
ACCESSIONS I54523
REFERENCE  I54523

```

```

#authors      Ando, A.; Yara-Kikuti, Y.; Kawata, H.; Okamoto, N.; Imai, T.;
               Eki, T.; Yokoyama, K.; Soeda, E.; Ikemura, T.; Abe, K.;
#journal      Immunogenetics (1994) 39:194-200
#title        Cloning of a new kinesin-related gene located at the
               centromeric end of the human MHC region.
#cross-references MUID:94102819
#accession    154523
#status       Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-519 ##label RES
##cross-references GB:D14678; NID:9510281; PID:g510282
CLASSIFICATION #superfamily unassigned kinesin-related proteins; kinesin
               motor domain homology
KEYWORDS      ATP; P-loop
FEATURE       157-515
               #domain kinesin motor domain homology #label KMOT\
256-263        #region nucleotide-binding motif A (P-loop)
SUMMARY       #length 519 #checksum 4377

Query Match      65.7%; Score 44; DB 2; Length 519;
Best Local Similarity 66.7%; Pred. NO. 1.93e-01;
Matches          6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 269 GPGDPQLE 277
|||:||||
Qy 2 GPTGEPOQE 10

RESULT 13
ENTRY   S30059 #type complete
TITLE   Probable transport protein ALD - human
ALTERNATE_NAMES
ORGANISM #formal_name Homo sapiens #common_name man
DATE     31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
17-Mar-1999
ACCESSIONS S30059
REFERENCE   Mosser, J.; Douar, A.M.; Sarde, C.O.; Kioschis, P.; Feil, R.;
               Moser, H.; Poustka, A.M.; Mandel, J.L.; Aubourg, P.
#journal    Nature (1993) 361:726-730
#title      Putative X-linked adrenoleukodystrophy gene shares unexpected
               homology with ABC transporters.
#cross-references MUID:93180910
#accession  S30059
#molecule_type mRNA
#residues   1-745 ##label MOS
##cross-references EMBL:Z21876; NID:g38590; PID:g38591
GENETICS
#gene       GDB:ALD
#map_position Xq28-Xq28
#cross-references GDB:118991; OMIM:300100
CLASSIFICATION #superfamily ATP-binding cassette homology
               adrenoleukodystrophy; ATP; membrane protein; P-loop
FEATURE      490-575
               #domain ATP-binding cassette homology #label ABC\
507-514        #region nucleotide-binding motif A (P-loop)\
513            #binding site ATP (Lys) #status predicted
SUMMARY      #length 745 #molecular-weight 82908 #checksum 9250

Query Match      65.7%; Score 44; DB 2; Length 745;
Best Local Similarity 60.0%; Pred. NO. 1.93e-01;
Matches          6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPTQE 56
|||:||||
Qy 1 DGPTGEPOQE 10

RESULT 14
ENTRY   G02500 #type complete
TITLE   adrenoleukodystrophy protein - human
ORGANISM #formal_name Homo sapiens #common_name man

#authors      Ando, A.; Yara-Kikuti, Y.; Kawata, H.; Okamoto, N.; Imai, T.;
               Eki, T.; Yokoyama, K.; Soeda, E.; Ikemura, T.; Abe, K.;
#journal      Immunogenetics (1994) 39:194-200
#title        Cloning of a new kinesin-related gene located at the
               centromeric end of the human MHC region.
#cross-references MUID:94102819
#accession    154523
#status       Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-519 ##label RES
##cross-references GB:D14678; NID:9510281; PID:g510282
CLASSIFICATION #superfamily unassigned kinesin-related proteins; kinesin
               motor domain homology
KEYWORDS      ATP; P-loop
FEATURE       157-515
               #domain kinesin motor domain homology #label KMOT\
256-263        #region nucleotide-binding motif A (P-loop)
SUMMARY       #length 519 #checksum 4377

Query Match      65.7%; Score 44; DB 2; Length 519;
Best Local Similarity 66.7%; Pred. NO. 1.93e-01;
Matches          6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 269 GPGDPQLE 277
|||:||||
Qy 2 GPTGEPOQE 10

RESULT 13
ENTRY   S30059 #type complete
TITLE   Probable transport protein ALD - human
ALTERNATE_NAMES
ORGANISM #formal_name Homo sapiens #common_name man
DATE     31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
17-Mar-1999
ACCESSIONS S30059
REFERENCE   Mosser, J.; Douar, A.M.; Sarde, C.O.; Kioschis, P.; Feil, R.;
               Moser, H.; Poustka, A.M.; Mandel, J.L.; Aubourg, P.
#journal    Nature (1993) 361:726-730
#title      Putative X-linked adrenoleukodystrophy gene shares unexpected
               homology with ABC transporters.
#cross-references MUID:93180910
#accession  S30059
#molecule_type mRNA
#residues   1-745 ##label MOS
##cross-references EMBL:Z21876; NID:g38590; PID:g38591
GENETICS
#gene       GDB:ALD
#map_position Xq28-Xq28
#cross-references GDB:118991; OMIM:300100
CLASSIFICATION #superfamily ATP-binding cassette homology
               adrenoleukodystrophy; ATP; membrane protein; P-loop
FEATURE      490-575
               #domain ATP-binding cassette homology #label ABC\
507-514        #region nucleotide-binding motif A (P-loop)\
513            #binding site ATP (Lys) #status predicted
SUMMARY      #length 745 #molecular-weight 82908 #checksum 9250

Query Match      65.7%; Score 44; DB 2; Length 745;
Best Local Similarity 60.0%; Pred. NO. 1.93e-01;
Matches          6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPTQE 56
|||:||||
Qy 1 DGPTGEPOQE 10

RESULT 14
ENTRY   G02500 #type complete
TITLE   adrenoleukodystrophy protein - human
ORGANISM #formal_name Homo sapiens #common_name man

```

```

DATE          21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
18-Sep-1998
ACCESSIONS    G02500
REFERENCE      H01367
#authors      Platzer, M.; Bauer, D.; Drescher, B.
#submissions  submitted to the EMBL Data Library, March 1996
#accession    G02500
#status       Preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues     1-745 ##label PLA
##cross-references EMBL:U52111; NID:g1302649; PID:g1302652
GENETICS
#gene         GDB:ALD; AMN
#map_position Xq28-Xq28
#cross-references GDB:118991; OMIM:300100
#introns      300/3; 361/1; 408/3; 465/1; 496/3; 545/2; 594/1; 622/2; 664/2
CLASSIFICATION #superfamily ATP-binding cassette homology
FEATURE        490-675
               #domain ATP-binding cassette homology #label ABC\
507-514        #region nucleotide-binding motif A (P-loop)
SUMMARY       #length 745 #molecular-weight 82936 #checksum 9439

Query Match      65.7%; Score 44; DB 2; Length 745;
Best Local Similarity 60.0%; Pred. NO. 1.93e-01;
Matches          6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPTQE 56
|||:||||
Qy 1 DGPTGEPOQE 10

RESULT 15
ENTRY   S14113 #type complete
TITLE   1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase (EC
               3.1.4.11) delta-2 - bovine
ALTERNATE_NAMES inositol-phospholipid-specific phospholipase C
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE     21-Nov-1993 #sequence_revision 02-Jun-1995 #text_change
29-May-1998
ACCESSIONS S14113; S04944
REFERENCE   Meldrum, E.; Kriz, R.W.; Totty, N.; Parker, P.J.
               Eur. J. Biochem. (1991) 196:159-165
#journal     A second gene product of the inositol-phospholipid-specific
               phospholipase C-delta subclass.
#title       phospholipase C-delta subclass.
#cross-references MUID:91160548
#accession  S14113
#status      preliminary
#molecule_type mRNA
#residues    1-764 ##label MEL
#experimental_source brain
REFERENCE      S04944
#authors      Meldrum, E.; Katan, M.; Parker, P.
#journal      Eur. J. Biochem. (1989) 182:673-677
#title        A novel inositol-phospholipid-specific phospholipase C. Rapid
               purification and characterization.
#cross-references MUID:89325315
#accession  S04944
#status      preliminary
#molecule_type protein
#residues    528-541,'X',543-553;659-669 #label ME2
#experimental_source brain
CLASSIFICATION #superfamily 1-phosphatidylinositol-4,5-bisphosphate
               phosphodiesterase III; 1-phosphatidylinositol-4,
               5-bisphosphate phosphodiesterase domain X homology;
               1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase
               domain Y homology; pleckstrin repeat homology
               phosphoric diester hydrolase
KEYWORDS      phosphoric diester hydrolase
FEATURE       292-435
               #domain 1-phosphatidylinositol-4,5-bisphosphate
               phosphodiesterase domain X homology #label PIPX\
489-609        #domain 1-phosphatidylinositol-4,5-bisphosphate
               phosphodiesterase domain Y homology #label PIPY

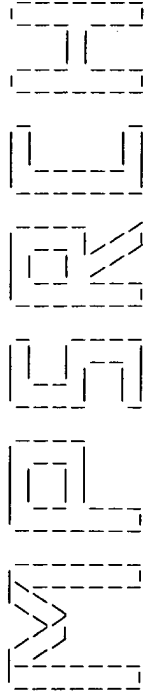
```

SUMMARY            #length 764   #molecular-weight 87681   #checksum 1361  
Query Match            65.7%;   Score 44;   DB 2;   Length 764;  
Best Local Similarity   85.7%;   Pred. No. 1.93e+01;  
Matches            6;   Conservative   1;   Mismatches   0;   Indels   0;   Gaps   0;  
  
Db   340   DGPSGEP 346  
         |||:||||  
Qy        1   DGPTGEP 7

Search completed: Thu Oct 21 15:42:12 1999  
Job time : 20 secs.



\*\*\*\*\*



\*\*\*\*\* (TM) \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPorch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 21 15:42:29 1999; MasPar time 2.16 Seconds  
Tabular output not generated. 130.846 Million cell updates/sec

Title: >US-09-040-485-8  
Description: (1-10) from US09040485.pep  
Perfect Score: 67  
Sequence: 1 DGPTGEPOQE 10

Scoring table: PAM 150  
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 21.166; Variance 23.548; scale 0.899

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	67	100.0	757	1	ASPH_HUMAN ASPARTYL/ASPARAGINYL B	1.35e-05
2	50	74.6	232	1	VG08_HSV11 HYPOTHETICAL GENE 8 ME	3.45e-01
3	47	70.1	433	1	TRBI_AGR16 CONJUGAL TRANSFER PROT	1.74e+00
4	47	70.1	1389	1	TIMDROME TIMELESS PROTEIN	1.74e+00
5	47	70.1	3567	1	ERY2_SAGER ERYTHRONOLIDE SYNTHASE	1.74e+00
6	46	68.7	311	1	SRY_MUSST SEX-DETERMINING REGION	2.95e+00
7	46	68.7	355	1	SRY_MUSST SEX-DETERMINING REGION	2.95e+00
8	45	67.2	53	1	CAL1_RABIT COLLAGEN ALPHA 1(I) CH	4.95e+00
9	45	67.2	259	1	YCB_CECOLI HYPOTHETICAL 28.7 KD P	4.95e+00
10	45	67.2	1027	1	P531_HUMAN P53-BINDING PROTEIN 53	4.95e+00
11	45	67.2	1418	1	CAL2_HUMAN PROCOLLAGEN ALPHA 1(II) 4.95e+00	
12	45	67.2	1596	1	NAM_DROME NEUROGENIC PROTEIN MAS	4.95e+00
13	44	65.7	261	1	PRP2_MOUSE PROLINE-RICH PROTEIN M	8.25e+00
14	44	65.7	745	1	ADL_HUMAN ADRENOLEUKODYSTROPHY P	8.25e+00
15	44	65.7	755	1	REEL_HUMAN RAS-RESPONSIVE ELEMENT	8.25e+00
16	44	65.7	1388	1	CAL1_HUMAN COLLAGEN ALPHA 1(XV) C	8.25e+00
17	43	64.2	179	1	SP17_MONDO SPERM SURFACE PROTEIN	1.36e+01
18	43	64.2	336	1	CAL6_BOVIN COLLAGEN ALPHA 1(VI) C	1.36e+01
19	43	64.2	449	1	CMGA_BOVIN CHROMOGGRANIN A PRECURS	1.36e+01
20	43	64.2	1670	1	CA34_HUMAN PROCOLLAGEN ALPHA 3(IV)	1.36e+01
21	43	64.2	1707	1	CA24_MOUSE PROCOLLAGEN ALPHA 2(IV)	1.36e+01
22	43	64.2	1841	1	RPB1_ARATH DNA-DIRECTED RNA POLYM	1.36e+01
23	43	64.2	1860	1	RPB0_ARATH DNA-DIRECTED RNA POLYM	1.36e+01

RESULT ID	ASPH_HUMAN	STANDARD	PRT	757 AA
AC	Q12797			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-DE DIOXYGENASE).			
GN	ASPH.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.			
RC	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 95121937.			
RA	KORIOTH F., GIEFFERS C., FREY J.:			
RT	"Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase."			
RL	GENE 150:395-399(1994).			
CC	-I- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.			
CC	-I- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) = PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).			
CC	-I- COFACTOR: IRON.			
CC	-I- SUBUNIT: MONOMER (BY SIMILARITY).			
CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.			
CC	-I- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.			
CC	-I- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: U03109; G458032; -			
DR	MM; 600582; -			
DR	OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;			
KW	ENDOPLASMIC RETICULUM.			
FT	DOMAIN 1 54			
				CYTOPLASMIC (POTENTIAL).

ALIGNMENTS

RESULT 1	ASPH_HUMAN	STANDARD	PRT	757 AA
AC	Q12797			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-DE DIOXYGENASE).			
GN	ASPH.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.			
RC	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 95121937.			
RA	KORIOTH F., GIEFFERS C., FREY J.:			
RT	"Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase."			
RL	GENE 150:395-399(1994).			
CC	-I- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.			
CC	-I- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) = PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).			
CC	-I- COFACTOR: IRON.			
CC	-I- SUBUNIT: MONOMER (BY SIMILARITY).			
CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.			
CC	-I- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.			
CC	-I- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: U03109; G458032; -			
DR	MM; 600582; -			
DR	OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;			
KW	ENDOPLASMIC RETICULUM.			
FT	DOMAIN 1 54			
				CYTOPLASMIC (POTENTIAL).

```

FT TRANSMEM 55 75 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 76 757 (POTENTIAL).
FT LUMENAL 13 20 LUMENAL (POTENTIAL).
FT POLY-SER. 323 332 POLY-SER.
FT POLY-LYS. 452 452 POLY-LYS.
FT CARBOHYD 705 705 POTENTIAL.
FT CARBOHYD 705 705 POTENTIAL.
SQ SEQUENCE 757 AA: 85498 MW; AEGAF24 CRC32;

Query Match 100.0%; Score 67; DB 1; Length 757;
Best Local Similarity 100.0%; Pred. No. 1.35e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 173 DGTGPEQOE 182
Qy 1 DGTGPEQOE 10

RESULT 2
ID VG08_HSV11 STANDARD; PRT; 232 AA.
AC Q00137;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL GENE 8 MEMBRANE PROTEIN.
GN 8.
OS ICTALURID HERPESVIRUS 1 (CHANNEL CATFISH VIRUS) (CCV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC UNCLASSIFIED HERPESVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AUBURN 1.
RX MEDLINE; 92087490.
RA DAVISON A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL VIROLOGY 186:9-14(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M75136; G331295; -
CC DR EMBL; M75136; G331218; -
CC DR PIR; I36786; MMBE13.
CC KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 166 182 POTENTIAL.
FT TRANSMEM 195 211 POTENTIAL.
SQ SEQUENCE 232 AA; 25466 MW; AA90E31E CRC32;

Query Match 74.6%; Score 50; DB 1; Length 232;
Best Local Similarity 50.0%; Pred. No. 3.45e-01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 88 ESPGTPHPRD 97
Qy 1 DGTGPEQOE 10

RESULT 3
ID TRBI_AGR6 STANDARD; PRT; 433 AA.
AC P54917;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CONJUGAL TRANSFER PROTEIN TRBI.
GN TRBI.
OS AGROBACTERIUM TUMEFACIENS.
OC PLASMID PTIASCN.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;

```

```

OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96312368.
RA ALT-MORBE J., STRYKER J.L., FUQUA C., LI P.L., FARRAND S.K.,
RA WINANS S.C.;
RT "The conjugal transfer system of Agrobacterium tumefaciens
RT octopine-type Ti plasmids is closely related to the transfer system
RT of an IncP plasmid and distantly related to Ti plasmid vir genes.";
RL J. BACTERIOL. 178:4248-4257(1996).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U43675; G2749901; -
CC DR CONJUGATION; PLASMID; TRANSMEMBRANE.
KW TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 307 327 POTENTIAL.
FT TRANSMEM 347 367 POTENTIAL.
SQ SEQUENCE 433 AA; 47258 MW; E3C35AE6 CRC32;

Query Match 70.1%; Score 47; DB 1; Length 433;
Best Local Similarity 70.0%; Pred. No. 1.74e-00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 79 DGIIGPEQOO 88
Qy 1 DGTGPEQOE 10

RESULT 4
ID TIM_DROME STANDARD; PRT; 1389 AA.
AC P49021;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE TIMELESS PROTEIN.
GN TIM.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96055118.
RA MYERS M.P., WAGER-SMITH K., WESLEY C.S., YOUNG M.W., SEHGAL A.;
RT "Positional cloning and sequence analysis of the Drosophila clock
RT gene, timeless.";
RL SCIENCE 270:805-808(1995).
CC [2]
CC INTERACTION WITH PER.
CC RX MEDLINE; 96055120.
CC RA GEKAKIS N., SAEZ L., DELAHAYE-BROWN A.M., MYERS M.P., SEHGAL A.,
CC YOUNG M.W., WEITZ C.J.;
CC "Isolation of timeless by PER protein interaction: defective
CC interaction between timeless protein and long-period mutant PERL.";
CC RL SCIENCE 270:811-815(1995).
CC -!- FUNCTION: REQUIRED FOR THE PRODUCTION OF CIRCADIAN RHYTHMS.
CC INTERACTS WITH PERIOD (PER). MAY BE REQUIRED AT A SPECIFIC
CC TIME OF DAY TO ALLOW ACCUMULATION AND NUCLEAR LOCALIZATION OF
CC THE PER PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U37018; G1050970; -.  
 DR FLYBASE: FBgn0014396; tim.  
 KW BIOLOGICAL RHYTHMS. 380 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 351 380 ARG/LYS-RICH (BASIC).  
 FT DOMAIN 540 553  
 SQ SEQUENCE 1389 AA; 155665 MW; 1DB78941 CRC32;

Query Match 70.1%; Score 47; DB 1; Length 1389;  
 Best Local Similarity 60.0%; Pred. No. 1.74e+00;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 DGPQGRPHQ 519  
 ||| | ||:  
 QY 1 DGPTEPQ 10

RESULT 5  
 ID ERY2\_SACER STANDARD; PRT; 3567 AA.  
 AC Q03132; Q54096;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE ERYTHRONOLIDE SYNTHASE, MODULES 3 AND 4 (EC 2.3.1.94) (ORF 2) (6-  
 DE DEOXYERYTHRONOLIDE B SYNTHASE II) (DEBS 2).  
 GN ERYA.  
 OS SACCCHAROPOLYSPORA ERYTHRAEA (STREPTOMYCES ERYTHRAEUS).  
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
 OC ACTINOMYCETALES; PSEUDONOCARDINEAE; PSEUDONOCARDIACEAE;  
 OC SACCCHAROPOLYSPORA.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91220065.  
 RA DONADIO S., STAVIER M.J., MCALPINE J.B., SWANSON S.J., KATZ L.;  
 RT "Modular organization of genes required for complex polyketide  
 RT biosynthesis.";  
 RL SCIENCE 252:675-679(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL 2338;  
 RX MEDLINE: 92155230.  
 RA BEVITT D.J., CORTES J., HAYDOCK S.F., LEADLAY P.F.;  
 RT "6-Deoxyerythronolide-B synthase 2 from Sacccharopolyspora erythraea.  
 RT Cloning of the structural gene, sequence analysis and inferred domain  
 RT structure of the multifunctional enzyme.";  
 RL EUR. J. BIOCHEM. 204:39-49(1992).  
 CC -!- CATALYTIC ACTIVITY: 6 METHYLMALONYL-COA + PROPIONYL-COA = 7 COA  
 CC + 6-DEOXYERYTHRONOLIDE B.  
 CC -!- COFACTOR: NADP.  
 CC -!- COFACTOR: CONTAINS TWO COVALENTLY BOUND PHOSPHOPANTETHEINES.  
 CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN  
 CC BIOSYNTHESIS.  
 CC -!- IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH ENCODING FOR A  
 CC FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3 ORFS CODES  
 CC FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH SYNTHASE  
 CC PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS  
 CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,  
 CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,  
 CC RESPECTIVELY.  
 CC -!- BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT), BETA-KETOACYL  
 CC CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER PROTEIN (ACP) FOR  
 CC CHAIN ELONGATION. BETA-KETOREDUCTASE (KR), DEHYDRATASE (DH), AND  
 CC ENOYL REDUCTASE (ER) FOR PROCESSING OF THE BETA CARBON. AND  
 CC THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION OF THE FULL-  
 CC LENGTH CHAIN.  
 CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M63677; G152694; -.  
 DR EMBL; X62569; G581651; -.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.  
 DR PROSITE; PS00606; B-KETOACYL SYNTHASE; 2.  
 DR PROSITE; PS00075; ACP\_DOMAIN; 2.  
 DR PFAM; PF00106; adh\_short; 1.  
 DR PFAM; PF00107; adh\_zinc; 1.  
 DR PFAM; PF00109; ketoacyl-synt; 2.  
 DR PFAM; PF00550; pp-binding; 2.  
 DR PFAM; PF00698; Acyl\_transf; 2.  
 DR HSSP; P28304; 10OR.  
 KW TRANSFERASE; ACYLTRANSFERASE; ANTIBIOTIC BIOSYNTHESIS; NADP;  
 KW PHOSPHOPANTETHEINE; MULTIFUNCTIONAL ENZYME.  
 FT DOMAIN 1 1484  
 FT DOMAIN 1485 3567  
 FT DOMAIN 27 488  
 FT DOMAIN 559 884  
 FT DOMAIN 1130 1301  
 FT DOMAIN 1397 1467  
 FT DOMAIN 1485 1943  
 FT DOMAIN 2013 2336  
 FT DOMAIN 2383 3066  
 FT DOMAIN 3139 3322  
 FT DOMAIN 3415 3485  
 FT ACT\_SITE 202 202  
 FT ACT\_SITE 651 651  
 FT BINDING 1430 1430  
 FT ACT\_SITE 1661 1661  
 FT ACT\_SITE 2115 2115  
 FT NP\_BIND 2961 2978  
 FT NP\_BIND 3142 3157  
 FT BINDING 3448 3448  
 FT CONFLICT 438 438  
 FT CONFLICT 480 480  
 FT CONFLICT 1241 1241  
 FT CONFLICT 2664 2664  
 SQ SEQUENCE 3567 AA; 374413 MW; A2F5EA2C CRC32;

Query Match 70.1%; Score 47; DB 1; Length 3567;  
 Best Local Similarity 87.5%; Pred. No. 1.74e+00;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1376 GPTEPQ 1383  
 ||||| |  
 QY 2 GPTEPQ 9

RESULT 6  
 ID SRY\_MUSSI STANDARD; PRT; 311 AA.  
 AC Q62565;  
 DT 15-JUL-1998 (REL. 36, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE SEX-DETERMINING REGION Y PROTEIN (TESTIS-DETERMINING FACTOR).  
 GN SRY OR TDY  
 OS MUS SPICILEGUS (STEPPE MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 98043417.  
 RA ALBRECHT K.H., EICHER E.M.;  
 RT "DNA sequence analysis of Sry alleles (subgenus Mus) implicates  
 RT misregulation as the cause of C57BL/6J-Y(pos) sex reversal and  
 RT defines the SRY functional unit.";  
 RL GENETICS 147:1267-1277(1997).  
 RN [2]  
 RP SEQUENCE OF 1-143 FROM N.A.

RX MEDLINE: 94285818.  
 RA LUNDRIGAN B.L., TUCKER P.K.;  
 RT "Tracing paternal ancestry in mice, using the Y-linked,  
 RL sex-determining locus, Sry.";  
 CC MOL. BIOL. EVOL. 11:483-492(1994).  
 CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC  
 CC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE  
 CC SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL  
 CC INTERCALATION IN THE MINOR GROOVE.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- DOMAIN: THE GLN- AND HIS-RICH DOMAIN MAY MEDIATE PROTEIN-PROTEIN  
 CC INTERACTIONS.  
 CC -!- SIMILARITY: CONTAINS 1 HMG BOX.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U70658; G2623381; -  
 CC EMBL: AF009520; G2271481; -  
 CC EMBL: L29550; G463146; -  
 CC MGD: MGI:98660; TDY.  
 CC PFAM: PF00505; HMG\_box; 1.  
 CC HSSP: Q05066; 1HRZ.  
 CC DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC SEXUAL DIFFERENTIATION; REPEAT.  
 CC DNA\_BIND 5 73 HMG BOX.  
 CC SEQUENCE 311 AA; 38469 MW; 035872C0 CRC32;  
 FT  
 SQ  
 Query Match 68.7%; Score 46; DB 1; Length 311;  
 Best Local Similarity 70.0%; Pred. No. 2.95e+00;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 137 DIPTGHPOQ 146  
 QY 1 DGPTGEPOQ 10

RESULT 7  
 ID SRY\_MUSSP STANDARD; PRT; 355 AA.  
 AC Q62563;  
 DT 15-JUL-1998 (REL. 36, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE SEX-DETERMINING REGION Y PROTEIN (TESTIS-DETERMINING FACTOR).  
 GN SRY.  
 OS MUS SPRETUS (WESTERN WILD MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 98043417.  
 RA ALBRECHT K.H., EICHER E.M.;  
 RT "DNA sequence analysis of Sry alleles (subgenus Mus) implicates  
 RT misregulation as the cause of C57BL/6J-Y(POS) sex reversal and  
 RT defines the SRY functional unit.";  
 RL GENETICS 147:1267-1277(1997).  
 RN [2]  
 RN SEQUENCE OF 1-143 FROM N.A.  
 RP MEDLINE: 93361118.  
 RX TUCKER P.K., LUNDRIGAN B.L.;  
 RT "Rapid evolution of the sex determining locus in Old world mice and  
 RT rats.";  
 RL NATURE 364:715-717(1993).  
 CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC  
 CC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE  
 CC SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL  
 CC INTERCALATION IN THE MINOR GROOVE.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- DOMAIN: THE GLN- AND HIS-RICH DOMAIN MAY MEDIATE PROTEIN-PROTEIN  
 CC INTERACTIONS.  
 CC -!- SIMILARITY: CONTAINS 1 HMG BOX.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U70659; G2623383; -  
 CC EMBL: AF009521; G2271483; -  
 CC EMBL: L29544; G496150; -  
 CC MGD: MGI:98660; TDY.  
 CC PFAM: PF00505; HMG\_box; 1.  
 CC HSSP: Q05066; 1HRZ.  
 CC DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR;  
 CC SEXUAL DIFFERENTIATION; REPEAT.  
 CC DNA\_BIND 5 73 HMG BOX.  
 CC SEQUENCE 355 AA; 44260 MW; 09C5ACAE CRC32;  
 FT  
 SQ  
 Query Match 68.7%; Score 46; DB 1; Length 355;  
 Best Local Similarity 70.0%; Pred. No. 2.95e+00;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Db 137 DIPTGHPOQ 146  
 QY 1 DGPTGEPOQ 10

RESULT 8  
 ID CALL\_RABIT STANDARD; PRT; 53 AA.  
 AC P02456;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE COLLAGEN ALPHA 1(I) CHAIN (FRAGMENT).  
 GN COL1A1.  
 OS ORYCTOLAGUS CUNICULUS (RABBIT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE: 70252720.  
 RA BORNSTEIN P., NESSE R.;  
 RT "The comparative biochemistry of collagen: the structure of rabbit  
 RT skin collagen and its relevance to immunochemical studies of  
 RT collagen.";  
 RL ARCH. BIOCHEM. BIOPHYS. 138:443-450(1970).  
 CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 CC (FIBRILLAR FORMING COLLAGEN).  
 CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 CC HYDROXYAPATITE.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC PIR: A02856; CGBR15.  
 DR PROSITE; PS01208; VWFCD.  
 KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;  
 KW COLLAGEN.  
 FT MOD\_RES 7 7 CONVERTED TO AN ALDEHYDE GROUP THAT IS  
 FT INVOLVED IN CROSS-LINKING.  
 FT MOD\_RES 26 26 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 29 29 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 32 32 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 41 41 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 44 44 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 47 47 HYDROXYLATION (PROBABLE).  
 FT NON\_TER 53 53

```
SQ SEQUENCE 53 AA; 4987 MW; B20D776E CRC32;
Query Match 67.2%; Score 45; DB 1; Length 53;
Best Local Similarity 44.4%; Pred. No. 4.95e+00;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 38 ZGPPGZPGZ 46
:|:|:|:
QY 1 DGPTGEPOQ 9

RESULT 9
ID YCBC_ECOLI STANDARD; PRT; 259 AA.
AC P36565; P75846;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 28.7 KD PROTEIN IN KDSB-MUKF INTERGENIC REGION.
GN YCBC.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGL1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL SCIENCE 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 1-170 FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE; 94232180.
RA FENG J., YAMANAKA K., NIKI H., OGURA T., HIRAGA S.;
RT "New killing system controlled by two genes located immediately
RT upstream of the mukB gene in Escherichia coli."
RL MOL. GEN. GENET. 243:136-147(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A5000194; G1787150;
DR EMBL; D26440; ; NOT_ANNOTATED_CDS.
DR ECOGENE; EG12166; YCBC.
KW HYPOTHETICAL PROTEIN.
FT CONFLICT 89 89
FT CONFLICT 150 171
FT CONFLICT 150 171
SQ SEQUENCE 259 AA; 28666 MW; CB9EFE64 CRC32;
Query Match 67.2%; Score 45; DB 1; Length 259;
Best Local Similarity 66.7%; Pred. No. 4.95e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 251 GSGGEPOQ 259
:|:|:|:|:
QY 2 GPTGEPOQ 10

RESULT 10
ID P531_HUMAN STANDARD; PRT; 1027 AA.
AC Q12888;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

SQ SEQUENCE 53BP1 (FRAGMENT).
GN TP53BP1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94286584.
RA IWABUCHI K., BARTEL P.L., LI B., MARRACCINO R., FIELDS S.;
RT "Two cellular proteins that bind to wild-type but not mutant p53."
RL PROC. NATL. ACAD. SCI. U.S.A. 91:6098-6102(1994).
CC -!- FUNCTION: BINDS TO THE CENTRAL DOMAIN OF P53.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09477; G488592;
FT NON_TER 1
FT DOMAIN 697 701 POLY-SER.
FT DOMAIN 815 819 POLY-GLU.
SQ SEQUENCE 1027 AA; 111134 MW; 06D3FABB CRC32;
Query Match 67.2%; Score 45; DB 1; Length 1027;
Best Local Similarity 60.0%; Pred. No. 4.95e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 808 DGPTGSSEE 817
|||||:|:|
QY 1 DGPTGEPOQ 10

RESULT 11
ID CAL2_HUMAN STANDARD; PRT; 1418 AA.
AC P02458;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PROCOLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].
GN COL2A1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90067946.
RA SU M.W., LEE B., RAMIREZ F., MACHADO M., HORTON W.;
RT "Nucleotide sequence of the full length cDNA encoding for human type
RT II procollagen."
RL NUCLEIC ACIDS RES. 17:9473-9473(1989).
RN [2]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE; 87031574.
RA NUNEZ A.M., KOHNO K., MARTIN G.R., YAMADA Y.;
RT "Promoter region of the human pro-alpha 1(II)-collagen gene."
RL GENE 44:11-16(1986).
RN [3]
RP SEQUENCE OF 432-1145 FROM N.A.
RA RAMIREZ F.;
RL SUBMITTED (DEC-1988) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 963-1418 FROM N.A.
RX MEDLINE; 85190534.
RA CHEAH K.S.E., STOKER N.G., GRIFFIN J.R., GROSVELD F.G., SOLOMON E.;
RT "Identification and characterization of the human type II collagen
RT gene (COL2A1)."
RL PROC. NATL. ACAD. SCI. U.S.A. 82:2555-2559(1985).
RN [5]
RP SEQUENCE OF 1120-1398 FROM N.A.
```

RT MEDLINE: 85306861.  
RA ELIMA K., MAEKELAE J.K., VUORIO T., KAUPPINEN S., KNOWLES J.,  
RT VUORIO E.;  
RT "Construction and identification of a cDNA clone for human type II  
RT procollagen mRNA.";  
RL BIOCHEM. J. 229:183-188(1985).  
FN [6]  
RP SEQUENCE OF 1106-1418 FROM N.A.  
RX MEDLINE: 88067771.  
RA ELIMA K., VUORIO T., VUORIO E.;  
RT "Determination of the single polyadenylation site of the human pro  
RT alpha 1(II) collagen gene.";  
RL NUCLEIC ACIDS RES. 15:9499-9504(1987).  
FN [7]  
RP SEQUENCE OF 1227-1289 FROM N.A.  
RX MEDLINE: 86104135.  
RA NUNEZ A.M., FRANCOMANO C., YOUNG M.F., MARTIN G.R., YAMADA Y.;  
RT "Isolation and partial characterization of genomic clones coding for  
RT a human pro-alpha 1 (II) collagen chain and demonstration of  
RT restriction fragment length polymorphism at the 3' end of the gene.";  
RL BIOCHEMISTRY 24:6343-6348(1985).  
FN [8]  
RP SEQUENCE OF 1176-1226 FROM N.A.  
RX MEDLINE: 84118798.  
RA STROM C.M., UPHOLT W.B.;  
RT "Isolation and characterization of genomic clones corresponding to  
RT the human type II procollagen gene.";  
RL NUCLEIC ACIDS RES. 12:1025-1038(1984).  
FN [9]  
RP SEQUENCE OF 35-167 FROM N.A.  
RX MEDLINE: 89233138.  
RA SU M.W., BENSON-CHANDA V., VISSING H., RAMIREZ F.;  
RT "Organization of the exons coding for pro alpha 1(II) collagen N-  
RT propeptide confirms a distinct evolutionary history of this domain of  
RT the fibrillar collagen genes.";  
RL GENOMICS 4:438-441(1989).  
FN [10]  
RP REVIEW ON VARIANTS.  
RX MEDLINE: 91184577.  
RA KUIVANEMI H., TROMP G., PROCKOP D.J.;  
RT "Mutations in collagen genes: causes of rare and some common diseases  
RT in humans.";  
RL FASEB J. 5:2052-2060(1991).  
FN [11]  
RP REVIEW ON VARIANTS.  
RX MEDLINE: 97255959.  
RA KUIVANEMI H., TROMP G., PROCKOP D.J.;  
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
RT associated collagen (type IX), and network-forming collagen (type X)  
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
RL HUM. MUTAT. 9:300-315(1997).  
FN [12]  
RP VARIANT SER-1074.  
RX MEDLINE: 90036909.  
RA VISSING H., D'ALESSIO M., LEE B., RAMIREZ F., GODFREY M.,  
RA HOLLISTER D.W.;  
RT "Glycine to serine substitution in the triple helical domain of pro-  
RT alpha 1 (II) collagen results in a lethal perinatal form of short-  
RT limbed dwarfism.";  
RL J. BIOL. CHEM. 264:18265-18267(1989).  
FN [13]  
RP VARIANT SEDC GLY-1095--TYR-1330 DEL.  
RX MEDLINE: 89266907.  
RA LEE B., VISSING H., RAMIREZ F., ROGERS D., RIMOIN D.;  
RT "Identification of the molecular defect in a family with  
RT spondyloepiphyseal dysplasia.";  
RL SCIENCE 244:978-980(1989).  
FN [14]  
RP VARIANT OSTEOARTHRTIS CYS-650.  
RX MEDLINE: 90370826.  
RA ALA-KOKKO L., BALDWIN C.T., MOSKOWITZ R.W., PROCKOP D.J.;  
RT "Single base mutation in the type II procollagen gene (COL2A1) as a  
RT cause of primary osteoarthritis associated with a mild

RT chondrodysplasia.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 87:6565-6568(1990).  
FN [15]  
RP VARIANT OI-IV VAL-717.  
RX MEDLINE: 91291136.  
RA BAYEMAN J.F., HANNAGAN M., CHAN D., COLE W.G.;  
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to  
RT valine substitution in osteogenesis imperfecta type IV. Detection of  
RT the mutation and prenatal diagnosis by a chemical cleavage method.";  
RL BIOCHEM. J. 276:765-770(1991).  
FN [16]  
RP VARIANT OSTEOARTHRTIS CYS-650.  
RX MEDLINE: 91086471.  
RA EYRE D.R., WEIS M.A., MOSKOWITZ R.W.;  
RT "Cartilage expression of a type II collagen mutation in an inherited  
RT form of osteoarthritis associated with a mild chondrodysplasia.";  
RL J. CLIN. INVEST. 87:357-361(1991).  
FN [17]  
RP VARIANT HYPOCHONDROGENESIS GLU-984.  
RX MEDLINE: 93054548.  
RA BOGAERT R., TILLER G.E., WIES M.A., GRUBER H.E., RIMOIN D.L.,  
RA COHN D.H., EYRE D.R.;  
RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha  
RT 1(II) chain produces hypochondrogenesis.";  
RL J. BIOL. CHEM. 267:22522-22526(1992).  
FN [18]  
RP VARIANT HYPOCHONDROGENESIS SER-705.  
RX MEDLINE: 92262484.  
RA HORTON W.A., MACHADO M.A., ELLARD J., CAMPBELL D., BARTLEY J.,  
RA RAMIREZ F., VITALE E., LEE B.;  
RT "Characterization of a type II collagen gene (COL2A1) mutation  
RT identified in cultured chondrocytes from human hypochondrogenesis.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 89:4583-4587(1992).  
FN [19]  
RP VARIANT WS-II ASP-198.  
RX MEDLINE: 93304428.  
RA KORKKO J., RITVANEMI P., HAATAJA L., KAARIAINEN H., KIVIRIKKO K.I.,  
RA PROCKOP D.J., ALA-KOKKO L.;  
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate  
RT for glycine alpha 1-67 and that causes cataracts and retinal  
RT detachment: evidence for molecular heterogeneity in the Wagner  
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";  
RL AM. J. HUM. GENET. 53:55-61(1993).  
FN [20]  
RP VARIANT SEDC CYS-940.  
RX MEDLINE: 92262484.  
RA TILLER G.E., WEIS M.A., LACHMAN R.S., COHN D.H., RIMOIN D.L.,  
RA EYRE D.R.;  
RT "A dominant mutation in the type II collagen gene (COL2A1) produces  
RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";  
RL AM. J. HUM. GENET. 53:A209-A209(1993).  
FN [21]  
RP VARIANT OSTEOARTHRTIS CYS-650.  
RX MEDLINE: 93282819.  
RA HOLDERBAUM D., MALEUD C.J., MOSKOWITZ R.W., HAQOI T.M.;  
RT "Human cartilage from late stage familial osteoarthritis transcribes  
RT type II collagen mRNA encoding a cysteine in position 519.";  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 192:1169-1174(1993).  
FN [22]  
RP VARIANT SEDC ARG-285.  
RX MEDLINE: 93252400.  
RA VIRKULA M., RITVANEMI P., VUORIO A.F., KAITILÄ I., ALA-KOKKO L.,  
RA PELTONEN L.;  
RT "A mutation in the amino-terminal end of the triple helix of type II  
RT collagen causing severe osteochondrodysplasia.";  
RL GENOMICS 16:282-285(1993).  
FN [23]  
RP VARIANT SEDC CYS-206.  
RX MEDLINE: 94063862.  
RA WILLIAMS C.J., CONSIDINE E.L., KNOWLTON R.G., REGINATO A., NEUMANN G.,  
RA HARRISON D., BUXTON P., JIMENEZ S.A., PROCKOP D.J.;  
RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a  
RT family with an Arg75-->Cys mutation in the procollagen type II gene  
RT (COL2A1).";

```

RL HUM. GENET. 92:499-505(1993).
RN [24]
RP VARIANT SEDC CYS-920.
RX MEDLINE; 93315508.
RA CHAN D., TAYLOR T.K.F., COLE W.G.;
RT "Characterization of an arginine 789 to cysteine substitution in
RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
RT dysplasia.";
RL J. BIOL. CHEM. 268:15238-15245(1993).
RN [25]
RP VARIANT SEDC SER-1128.
RX MEDLINE; 93140139.
RA COLE W.G., HALL R.K., ROGERS J.G.;
RT "The clinical features of spondyloepiphyseal dysplasia congenita
RT resulting from the substitution of glycine 997 by serine in the alpha
RT 1(II) chain of type II collagen.";
RL J. MED. GENET. 30:27-35(1993).
RN [26]
...: remainder of annotations omitted.

Query Match 67.2%; Score 45; DB 1; Length 1418;
Best Local Similarity 56.7%; Pred. No. 4.95e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 930 GPSGEPQQ 938
II:III I;
QY 2 PTGEPQQE 10

RESULT 12
ID MAM_DROME STANDARD; PRT; 1596 AA.
AC P21519;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE NEUROGENIC PROTEIN MASTERMIND.
GN MAM.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE; 91065516.
RA SMOLLER D., FRIEDEL C., SCHMID A., BETTLER D., LAM L.,
RA YEDVOBNICK B.;
RT "The Drosophila neurogenic locus mastermind encodes a nuclear protein
RT unusually rich in amino acid homopolymers.";
RL GENES DEV. 4:1688-1700(1990).
CC -!- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
CC WITH THE N GENE PRODUCT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
CC UBQUITOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
CC CENTRAL NERVOUS SYSTEM.
CC -!- THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21 POLY-GLN RUNS
CC (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10 AA), 3 POLY-ASN
CC (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5 AA) RUNS.
CC -!- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54251; G8204; -
CC PIR; A33106; A33106.

```

```

DR PIR; A36391;
DR FLYBASE; FBgn0002643; mam.
KW NEUROGENESIS; NUCLEAR PROTEIN; REPEAT.
FT DOMAIN 20 84
FT DOMAIN 127 190
FT DOMAIN 196 219
FT DOMAIN 259 304
FT DOMAIN 355 388
FT DOMAIN 392 406
FT DOMAIN 407 440
FT DOMAIN 651 671
FT DOMAIN 700 714
FT DOMAIN 759 816
FT DOMAIN 987 996
FT DOMAIN 1060 1079
FT DOMAIN 1092 1107
FT DOMAIN 1237 1252
FT DOMAIN 1492 1496
FT DOMAIN 1559 1592
SQ SEQUENCE 1596 AA; 167717 MW; BEBC0500 CRC32;

Query Match 67.2%; Score 45; DB 1; Length 1596;
Best Local Similarity 66.7%; Pred. No. 4.95e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1317 GPMGGPQQ 1325
II:III I;
QY 2 PTGEPQQE 10

RESULT 13
ID PRP2_MOUSE STANDARD; PRT; 261 AA.
AC P05142;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROLINE-RICH PROTEIN MP-2 PRECURSOR.
GN PRP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86059475.
RA ANN D.K., CARLSON D.M.;
RT "The structure and organization of a proline-rich protein gene of a
RT mouse multigene family.";
RL J. BIOL. CHEM. 260:15863-15872(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M12099; G200547; -
CC REPEAT; SALIVA; SIGNAL.
FT SIGNAL 1 15
FT CHAIN 16 261
SQ SEQUENCE 261 AA; 26034 MW; 9D830DAF CRC32;

Query Match 65.7%; Score 44; DB 1; Length 261;
Best Local Similarity 75.0%; Pred. No. 8.25e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 193 PTGPGPQQ 200
II:III I;
QY 3 PTGEPQQE 10

RESULT 14

```

ID AC P33897; STANDARD; PRT; 745 AA.  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE ADRENOLEUKODYSTROPHY PROTEIN (ALDP).  
GN ALD.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93180910.  
RA MOSSER J., DOUAR A.-M., SARDE C.-O., KIOSCHIS P., FEIL R., MOSER H.,  
RA POUSTKA A.-M., MANDEL J.-L., AUBOURG P.;  
RT "Putative X-linked adrenoleukodystrophy gene shares unexpected  
RT homology with ABC transporters.";  
RL NATURE 361:726-730(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA PLATZER M., BAUER D., BRENNER V., DRESCHER B., NYAKATURA G.,  
RA REICHWALD K., SANDOVAL N., COY J., KIOSCHIS P., KORN B.,  
RA POUSTKA A.-M., ROSENTHAL A.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 97338663.  
RA DODD A., ROWLAND S.A., HAWKES S.L.J., KENNEDY M.A., LOVE D.R.;  
RT "Mutations in the adrenoleukodystrophy gene.";  
RL HUM. MUTAT. 9:500-511(1997).  
RN [4]  
RP VARIANT X-ALD LYS-291.  
RX MEDLINE; 94108454.  
RA CARTIER N., SARDE C.-O., DOUAR A.-M., MOSSER J., MANDEL J.-L.,  
RA AUBOURG P.;  
RT "Abnormal messenger RNA expression and a missense mutation in  
RT patients with X-linked adrenoleukodystrophy.";  
RL HUM. MOL. GENET. 2:1949-1951(1993).  
RN [5]  
RP REVIEW.  
RX MEDLINE; 93283453.  
RA AUBOURG P., MOSSER J., DOUAR A.-M., SARDE C.-O., LOPEZ J.,  
RA MANDEL J.-L.;  
RT "Adrenoleukodystrophy gene: unexpected homology to a protein involved  
RT in peroxisome biogenesis.";  
RL BIOCHIMIE 75:293-302(1993).  
RN [6]  
RP VARIANTS X-ALD SER-148; ASP-174; ARG-266; GLN-401; TRP-418 & PHE-515.  
RX MEDLINE; 95152524.  
RA FUCHS S., SARDE C.-O., WEDEMANN H., SCHWINGER E., MANDEL J.-L.,  
RA GAL A.;  
RT "Missense mutations are frequent in the gene for X-chromosomal  
RT adrenoleukodystrophy (ALD).";  
RL HUM. MOL. GENET. 3:1903-1905(1994).  
RN [7]  
RP VARIANTS X-ALD TRP-518; LEU-606; CYS-617 AND HIS-617.  
RX MEDLINE; 94314951.  
RA FANEN P., GUIDOUX S., SARDE C.-O., MANDEL J.-L., GOOSSENS M.,  
RA AUBOURG P.;  
RT "Identification of mutations in the putative ATP-binding domain of  
RT the adrenoleukodystrophy gene.";  
RL J. CLIN. INVEST. 94:516-520(1994).  
RN [8]  
RP VARIANTS X-ALD.  
RX MEDLINE; 95126139.  
RA LIGHTENBERG M.J.L., KEMP S., SARDE C.-O., VAN GEEL B.M., KLEIJER W.J.,  
RA BARTH P.G., MANDEL J.-L., VAN OOST B.A., BOLHUIS P.A.;  
RT "Spectrum of mutations in the gene encoding the adrenoleukodystrophy  
RT protein.";  
RL AM. J. HUM. GENET. 56:44-50(1995).  
RN [9]  
RP VARIANTS X-ALD HIS-104; GLU-178; LEU-560 AND GLY-528 DEL.  
RX MEDLINE; 95233433.

RA BRAUN A., AMBACH H., KAMMERER S., ROLINSKI B., STOECKLER S., RABL W.,  
RA GAETNER J., ZIERZ S., ROSCHER A.A.;  
RT "Mutations in the gene for X-linked adrenoleukodystrophy in patients  
RT with different clinical phenotypes.";  
RL AM. J. HUM. GENET. 56:854-861(1995).  
RN [10]  
RP VARIANTS X-ALD.  
RX MEDLINE; 96047143.  
RA KOK F., NEUMANN S., SARDE C.-O., ZHENG S., WU K.-H., WEI H.-M.,  
RA BERGIN J., WATKINS P.A., GOULD S., SACK G., MOSER H., MANDEL J.-L.,  
RA SMITH K.D.;  
RT "Mutational analysis of patients with X-linked adrenoleukodystrophy.";  
RL HUM. MUTAT. 6:104-115(1995).  
RN [11]  
RP VARIANTS X-ALD.  
RX MEDLINE; 96213748.  
RA FEIGENBAUM V., LOMBARD-PLATET G., GUIDOUX S., SARDE C.-O.,  
RA MANDEL J.-L., AUBOURG P.;  
RT "Mutational and protein analysis of patients and heterozygous women  
RT with X-linked adrenoleukodystrophy.";  
RL AM. J. HUM. GENET. 58:1135-1144(1996).  
RN [12]  
RP VARIANTS X-ALD.  
RX MEDLINE; 96163493.  
RA KRAEMANN E.W., MEIER V., KORENKE G.C., HUNNEMAN D.H., HANEFELD F.;  
RT "Identification of mutations in the ALD-gene of 20 families with  
RT adrenoleukodystrophy/adrenomyeloneuropathy.";  
RL HUM. GENET. 97:194-197(1996).  
RN [13]  
RP VARIANT AMN ARG-679.  
RX MEDLINE; 98112466.  
RA KORENKE G.C., KRAEMANN E., MEIER V., BEUCHE W., HUNNEMAN D.H.,  
RA HANEFELD F.;  
RT "First missense mutation (W679R) in exon 10 of the  
RT adrenoleukodystrophy gene in sibs with adrenomyeloneuropathy.";  
RL HUM. MUTAT. SUPPL. 1:S204-S206(1998).  
CC -!- FUNCTION: PROBABLE TRANSPORTER. COULD BE INVOLVED IN THE  
CC IMPORT OF VLCFA-COA SYNTHETASE INTO THE PEROXISOMAL MEMBRANE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL  
CC (PROBABLE).  
CC  
CC -!- DISEASE: DEFECTS IN ALD ARE THE CAUSE OF RECESSIVE X-LINKED  
CC ADRENOLEUKODYSTROPHY (X-ALD), A RARE PEROXISOMAL METABOLIC  
CC DISORDER THAT OCCURS IN BOYS AND IS CHARACTERIZED BY PROGRESSIVE  
CC MULTIFOCAL DEMYELINATION OF THE CENTRAL NERVOUS SYSTEM AND BY  
CC ADRENOCORTICAL INSUFFICIENCY. IT PRODUCES MENTAL DETERIORATION,  
CC CORTICOSPINAL TRACT DYSFUNCTION, AND CORTICAL BLINDNESS. THERE IS  
CC LABORATORY EVIDENCE OF ADRENAL CORTICAL DYSFUNCTION. DEATH  
CC INvariably OCCURS IN 1 TO 5 YEARS. DIFFERENT CLINICAL  
CC MANIFESTATIONS EXIST LIKE: CEREBRAL CHILDHOOD ALD (CALD), ADULT  
CC CEREBRAL ALD (ACALD), ADRENOLEUKODYSTROPHY (AMN) AND "ADDISON  
CC DISEASE ONLY" (ADO) PHENOTYPE.  
CC  
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
CC (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.  
CC  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; 221876; G38591; -  
DR EMBL; 231348; G806557; -  
DR EMBL; 231006; G806557; JOINED.  
DR EMBL; 231007; G806557; JOINED.  
DR EMBL; 231008; G806557; JOINED.  
DR EMBL; 231009; G806557; JOINED.  
DR EMBL; 231010; G806557; JOINED.  
DR EMBL; U52111; G1302652; -  
DR PIR; S30059; S30059.  
DR MIM; 300100; -  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.



DR PFAM; PF00005; ABC\_tran; 1.  
 KW ATP-BINDING; GLYCOPROTEIN; TRANSMEMBRANE; TRANSPORT; PEROXISOME;  
 FT DISEASE MUTATION; POLYMORPHISM.  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT TRANSMEM 131 151 POTENTIAL.  
 FT TRANSMEM 238 258 POTENTIAL.  
 FT TRANSMEM 333 353 POTENTIAL.  
 FT TRANSMEM 473 493 POTENTIAL.  
 FT NP\_BIND 507 514 ATP (BY SIMILARITY).  
 FT VARIANT 98 98 S -> L (IN X-ALD; CALD TYPE).  
 FT VARIANT 104 104 R -> C (IN X-ALD).  
 FT VARIANT 104 104 R -> H (IN X-ALD; ADO-TYPE).  
 FT VARIANT 104 104 T -> I (IN X-ALD; ADO-TYPE).  
 FT VARIANT 105 105 L -> P (IN X-ALD; ALD/AMN/ADO-TYPES AND  
 FT VARIANT 107 107 ASYMPTOMATIC).  
 FT VARIANT 108 108 S -> W (IN X-ALD; CALD AND AMN-TYPES).  
 FT VARIANT 116 116 G -> R (IN X-ALD; CALD-TYPE).  
 FT VARIANT 123 123 A -> V.  
 FT VARIANT 138 141 MISSING (IN X-ALD; ALD-TYPE).  
 FT VARIANT 141 141 A -> T (IN X-ALD).  
 FT VARIANT 148 148 N -> S (IN X-ALD; ADO-TYPE).  
 FT VARIANT 149 149 S -> N (IN X-ALD).  
 FT VARIANT 152 152 R -> C (IN X-ALD; ADO-TYPE).  
 FT VARIANT 152 152 R -> P (IN X-ALD).  
 FT VARIANT 152 152 R -> H (IN X-ALD).  
 FT VARIANT 163 163 Y -> D (IN X-ALD; ALD-TYPE).  
 FT VARIANT 174 174 Y -> S (IN X-ALD; CALD-TYPE).  
 FT VARIANT 174 174 Q -> E (IN X-ALD; AMN-TYPE).  
 FT VARIANT 178 178 Y -> C (IN X-ALD; ALMD-TYPE).  
 FT VARIANT 181 181 R -> P (IN X-ALD).  
 FT VARIANT 182 182 D -> V (IN X-ALD; CALD-TYPE).  
 FT VARIANT 194 194 D -> H (IN X-ALD).  
 FT VARIANT 200 200 L -> P (IN X-ALD).  
 FT VARIANT 211 211 L -> P (IN X-ALD).  
 FT VARIANT 220 220 L -> G (IN X-ALD).  
 FT VARIANT 221 221 D -> P (IN X-ALD; CALD AND AMN-TYPES).  
 FT VARIANT 254 254 T -> M (IN X-ALD; AMN-TYPE).  
 FT VARIANT 254 254 T -> P (IN X-ALD; AMN-TYPE).  
 FT VARIANT 263 263 P -> L (IN X-ALD; CALD, AMN AND AD-  
 FT VARIANT 263 263 TYPES).  
 FT VARIANT 266 266 G -> R (IN X-ALD).  
 FT VARIANT 276 276 K -> E (IN X-ALD; CALD-TYPE).  
 FT VARIANT 277 277 G -> R (IN X-ALD; AMN-TYPE).  
 FT VARIANT 277 277 G -> GN (IN X-ALD; ADO-TYPE).  
 FT VARIANT 277 277 G -> W (IN X-ALD).  
 FT VARIANT 281 281 E -> D (IN X-ALD; ACALD AND CALD-TYPES).  
 FT VARIANT 291 291 E -> K (IN X-ALD).  
 FT VARIANT 291 291 MISSING (IN X-ALD; ALD-TYPE).  
 FT VARIANT 294 294 A -> T (IN X-ALD; ALD-TYPE).  
 FT VARIANT 342 342 S -> P (IN X-ALD; AMN-TYPE).  
 FT VARIANT 389 389 R -> G (IN X-ALD; AMN-TYPE).  
 FT VARIANT 389 389 R -> H (IN X-ALD).  
 FT VARIANT 401 401 R -> Q (IN X-ALD).  
 FT VARIANT 418 418 R -> W (IN X-ALD; AMN-TYPE).  
 FT VARIANT 484 484 P -> R (IN X-ALD; CALD, AMN AND

.... Note: remainder of annotations omitted.

Query Match 65.7%; Score 44; DB 1; Length 745;  
 Best Local Similarity 60.0%; Pred. No. 8.25e+00;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPQOE 56  
 :|:|:|:|:|  
 QY 1 DGPTGEPPQE 10

RESULT 15  
 ID RREI\_HUMAN STANDARD; PRT; 755 AA.  
 AC Q92766;  
 DT 15-JUL-1998 (REL. 36, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE RAS-RESPONSIVE ELEMENT BINDING PROTEIN 1 (RREB-1).

GN RREB1.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-THYROID CARCINOMA;  
 RC MEDLINE: 96413283.  
 RA THIAGALINGAM A., DE BUSTROS A., BORGES M., JASTI R., COMPTON D.,  
 RA "DIAMOND L., MABRY M., BALL D.W., BAYLIN S.B., NELKIN B.D.;  
 RT "RREB-1, a novel zinc finger protein, is involved in the  
 RT differentiation response to Ras in human medullary thyroid  
 RT carcinomas";  
 RL MOL. CELL. BIOL. 16:5335-5345(1996).  
 CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS SPECIFICALLY TO THE  
 CC DISTAL RAS-RESPONSIVE ELEMENT (RRE) IN THE CALCITONIN GENE  
 CC PROMOTER AND AUGMENT THE RAS/RAF-MEDIATED TRANSCRIPTIONAL RESPONSE  
 CC OF THAT PROMOTER. MAY BE INVOLVED IN RAS/RAF-MEDIATED CELL  
 CC DIFFERENTIATION.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, PLACENTA, LUNG, LIVER,  
 CC SKELETAL MUSCLE, KIDNEY AND PANCREAS. NOT FOUND IN THE BRAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U26914; G1654112; -  
 DR MIM: 602209; -  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 4.  
 DR PFAM: PF00096; zf-C2H2; 4.  
 DR HSSP: P08045; 1ZNF.  
 KW TRANSCRIPTION REGULATION; ACTIVATOR; ZINC-FINGER; METAL-BINDING;  
 KW DNA-BINDING; NUCLEAR PROTEIN.  
 FT ZN\_FING 315 337 C2H2-TYPE.  
 FT ZN\_FING 461 483 C2H2-TYPE.  
 FT ZN\_FING 580 602 C2H2-TYPE.  
 FT ZN\_FING 608 630 C2H2-TYPE.  
 SQ SEQUENCE 755 AA; 79865 MW; 50627236 CRC32;

Query Match 65.7%; Score 44; DB 1; Length 755;  
 Best Local Similarity 50.0%; Pred. No. 8.25e+00;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 501 EGPSAPEPQE 510  
 :|:|:|:|:|  
 QY 1 DGPTGEPPQE 10

Search completed: Thu Oct 21 15:42:37 1999  
 Job time : 8 secs.

\*\*\*\*\*  
M P S R P H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:42:54 1999; Maspar time 4.30 Seconds  
Tabular output not generated. 126.993 Million cell updates/sec

Title: >US-09-040-485-8  
Description: (1-10) from US09040485.pep  
Perfect Score: 67  
Sequence: 1 DGPTGEPPQE 10

Scoring table: PAM 150  
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sprenb19  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 20.498; Variance 23.611; scale 0.868

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	51	76.1	205	5	001718 23KDA HEAT SHOCK PROTE	3.75e-01
2	48	71.6	414	2	P73855 HYPOTHETICAL 44.8 KD P	1.90e+00
3	47	70.1	357	13	P73826 NEUROD.	3.21e+00
4	47	70.1	357	13	P79765 NEUROD-LIKE PROTEIN.	3.21e+00
5	47	70.1	676	5	O44431 CIRCADIAN CLOCK PROTEI	3.21e+00
6	47	70.1	1120	5	O20778 SIMILAR TO TRIPLE HELI	3.21e+00
7	47	70.1	1194	5	O44430 CIRCADIAN CLOCK PROTEI	3.21e+00
8	47	70.1	1343	5	O17482 TIMELESS (TM).	3.21e+00
9	47	70.1	1398	5	O44380 CIRCADIAN CLOCK PROTEI	3.21e+00
10	46	68.7	231	11	O35966 SEX DETERMINING PROTEI	5.40e+00
11	46	68.7	232	11	O35968 TESTIS DETERMINING-Y (	5.40e+00
12	46	68.7	233	11	O35860 SEX DETERMINING PROTEI	5.40e+00
13	46	68.7	234	11	O35859 SEX DETERMINING PROTEI	5.40e+00
14	46	68.7	234	11	O35858 SEX DETERMINING PROTEI	5.40e+00
15	46	68.7	238	2	O69004 FERRISIDEROPHORE RECP	5.40e+00
16	45	67.2	77	4	O75782 COL1A1 AND PDGFB FUSIO	9.03e+00
17	45	67.2	89	7	O19497 MHC CLASS II BETA 1 DO	9.03e+00
18	45	67.2	227	7	O31398 MHC CLASS II B-L BETA	9.03e+00
19	45	67.2	263	13	O73896 B LOCUS L BETA CHAIN 2	9.03e+00
20	45	67.2	263	7	Q31410 MHC CLASS II B-LB11-BE	9.03e+00

21	45	67.2	488	14	037935 POLYPROTEIN (FRAGMENT)	9.03e+00
22	45	67.2	720	1	073955 HYDROPEROXIDASE.	9.03e+00
23	45	67.2	805	4	O75410 TACCL.	9.03e+00
24	44	65.7	243	14	Q67788 26 KD PROTEIN.	1.50e+01
25	44	65.7	295	5	Q20927 F57B7.3 PROTEIN.	1.50e+01
26	44	65.7	306	5	Q20135 F38A3.1 PROTEIN.	1.50e+01
27	44	65.7	310	5	Q20282 SIMILAR TO CUTICLE COL	1.50e+01
28	44	65.7	389	5	Q19707 F22B5.3 PROTEIN.	1.50e+01
29	44	65.7	474	5	Q21975 SIMILAR TO C. ELGANS	1.50e+01
30	44	65.7	519	4	Q14834 KINESIN-RELATED PROTEI	1.50e+01
31	44	65.7	673	4	Q60887 CKCK0721Q.3 (KINESIN R	1.50e+01
32	44	65.7	924	2	P73172 CHEMOTAXIS PROTEIN CHE	1.50e+01
33	44	65.7	1110	13	Q91255 NF-180.	1.50e+01
34	44	65.7	1142	4	O14497 B120.	1.50e+01
35	44	65.7	1302	4	O14525 KIAA0289 (DJ423N22.1.1	1.50e+01
36	43	64.2	89	7	O19494 MHC CLASS II BETA 1 DO	1.50e+01
37	43	64.2	281	5	O44774 F33D11.3 PROTEIN.	2.46e+01
38	43	64.2	394	5	O46068 1-EVIDENCE-PREDICTED B	2.46e+01
39	43	64.2	441	10	O64927 STARCH SYNTHASE (FRAGM	2.46e+01
40	43	64.2	588	14	O65043 LMPI.	2.46e+01
41	43	64.2	609	11	O08671 KIFC1.	2.46e+01
42	43	64.2	867	10	O81210 RESPIRATORY BURST OXID	2.46e+01
43	43	64.2	944	10	O80342 ATRBOH F PROTEIN.	2.46e+01
44	43	64.2	944	10	O48538 RBOHAP108.	2.46e+01
45	43	64.2	948	10	O81211 RESPIRATORY BURST OXID	2.46e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	205 AA.
ID	001718			
AC	001718			
DT	01-JUL-1997	(TREMBLREL. 04, CREATED)		
DT	01-JUL-1997	(TREMBLREL. 04, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	23KDA HEAT SHOCK PROTEIN SCHSP23.			
OS	SARCOPHAGA GRASSIPALPIS.			
OC	EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;			
OC	PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; OESTROIDEA; SARCOPHAGIDAE;			
OC	SARCOPHAGA.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	YOCUM G.D., JOPLIN K.H., DENLINGER D.L.;			
RL	SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR	ENBL; U96099; G2058737; -			
DR	PFAM; PF00011; HSP20; 1.			
KW	HEAT SHOCK.			
SQ	SEQUENCE 205 AA; 22933 MW; 5164A349 CRC32;			

Query Match 76.1%; Score 51; DB 5; Length 205;  
Best Local Similarity 70.0%; Pred. No. 3.75e-01;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 194 DGANGEPPQE 203  
||: ||||:|  
QY 1 DGPTGEPPQE 10

RESULT 2  
ID P73855 PRELIMINARY; PRT; 414 AA.

AC P73855;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 44.8 KD PROTEIN.  
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).  
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;  
RA TABATA S.;  
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
[2]

RP SEQUENCE FROM N.A.  
RC STRAIN=PCC6803;  
RX MEDLINE: 97061201.  
RA KANEKO T., SAITO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
RA HOSOUCHI T., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
RA TABATA S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. PCC6803. II. Sequence determination of the entire  
RT genome and assignment of potential protein-coding regions.";  
RL DNA RES. 3:109-136(1996).  
DR EMBL: D90910; D1018647; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 414 AA; 44810 MW; 0E3A0E9E CRC32;

Query Match 71.6%; Score 48; DB 2; Length 414;  
Best Local Similarity 60.0%; Pred. No. 1.90e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 318 EGPTPEAAQQ 327  
:|||||:  
QY 1 DGPTGEPOQE 10

RESULT 3  
ID O73826 PRELIMINARY; PRT; 357 AA.  
AC O73826;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE NEUROD.  
GN NEUROD.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WHITE LECHORN;  
RA YAN R.-T., WANG S.-Z.;  
RL J. NEUROBIOL. 0:0-0(1998).  
DR EMBL: AF060885; G3094020; -.  
SQ SEQUENCE 357 AA; 38858 MW; 7D406FF3 CRC32;

Query Match 70.1%; Score 47; DB 13; Length 357;  
Best Local Similarity 66.7%; Pred. No. 3.21e+00;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 GPAGEPQAQ 17  
:|||||:  
QY 2 GPTGEPOQE 10

RESULT 4  
ID P79765 PRELIMINARY; PRT; 357 AA.  
AC P79765;  
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE NEUROD-LIKE PROTEIN.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WHITE LECHORN; TISSUE-RETINA;  
RX MEDLINE: 97454246.  
RA ROZTOCIL T., MATTER-SADZINSKI L., ALLIOD C., BALLIVET M.,  
RA MATTER J.M.;  
RT "NeuroM, a neural helix-loop-helix transcription factor, defines a  
RT new transition stage in neurogenesis.";  
RL DEVELOPMENT 124:3263-3272(1997).  
DR EMBL: Y09596; E283389; -.

DR PFAM: PF00010; HLH; 1.  
SQ SEQUENCE 357 AA; 38809 MW; 04A96823 CRC32;

Query Match 70.1%; Score 47; DB 13; Length 357;  
Best Local Similarity 66.7%; Pred. No. 3.21e+00;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 GPAGEPQAQ 17  
:|||||:  
QY 2 GPTGEPOQE 10

RESULT 5  
ID O44431 PRELIMINARY; PRT; 676 AA.  
AC O44431;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE CIRCADIAN CLOCK PROTEIN (FRAGMENT).  
GN TIM.  
OS DROSOPHILA HYDEI (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA OUSLEY A., ZAFARULLAH K., CHEN Y., EMERSON M., HICKMAN L., SEHGAL A.;  
RL GENETICS 0:0-0(1998).  
DR EMBL: AF038579; G2724130; -.  
FT NON\_TER 1  
ET NON\_TER 676  
SQ SEQUENCE 676 AA; 76022 MW; 83E96D47 CRC32;

Query Match 70.1%; Score 47; DB 5; Length 676;  
Best Local Similarity 60.0%; Pred. No. 3.21e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 357 DGPGKPKQH 366  
:|||||:  
QY 1 DGPTGEPOQE 10

RESULT 6  
ID Q20778 PRELIMINARY; PRT; 1120 AA.  
AC Q20778;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE SIMILAR TO TRIPLE HELICAL REGION OF COLLAGENS.  
GN F54D8.1.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KESHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 0:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA BENTLEY D.;  
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]

RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U12966; G529221; - - - - -  
SQ SEQUENCE 1120 AA; 119368 MW; E0352B8E CRC32;

Query Match 70.1%; Score 47; DB 5; Length 1120;  
Best Local Similarity 60.0%; Pred. No. 3.21e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1020 EGPTEGPGAD 1029  
:|||||:  
QY 1 DGPTGEPQOE 10

RESULT 7  
ID O4430 PRELIMINARY; PRT; 1194 AA.  
AC O4430;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE CIRCADIAN CLOCK PROTEIN (FRAGMENT).  
GN TIM.  
OS DROSOPHILA VIRILIS (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
[1]

RP SEQUENCE FROM N.A.  
RA OUSLEY A., ZAFARULLAH K., CHEN Y., EMERSON M., HICKMAN L., SEHGAL A.;  
RL GENETICS 0:0-0(1998).  
DR EMBL; AF038502; G2746733; - -  
FT NONTER 1194 1194  
SQ SEQUENCE 1194 AA; 134946 MW; F6C4345B CRC32;

Query Match 70.1%; Score 47; DB 5; Length 1194;  
Best Local Similarity 60.0%; Pred. No. 3.21e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 486 DGPGKPKQH 495  
:|||||:  
QY 1 DGPTGEPQOE 10

RESULT 8  
ID O17482 PRELIMINARY; PRT; 1343 AA.  
AC O17482;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE TIMELESS (TIM).  
GN TIM.  
OS DROSOPHILA VIRILIS (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
[1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 98033379.  
RA MYERS M.P., ROTHENFLUH A., CHANG M., YOUNG M.W.;  
RL "Comparison of chromosomal DNA composing timeless in Drosophila melanogaster and D. virilis suggests a new conserved structure for the TIMELESS protein.";  
RL NUCLEIC ACIDS RES. 25:4710-4714(1997).  
DR EMBL; AF032403; G2641617; - -  
DR EMBL; AF032402; G2641617; JOINED.

SQ SEQUENCE 1343 AA; 150942 MW; 77573CAC CRC32;

Query Match 70.1%; Score 47; DB 5; Length 1343;  
Best Local Similarity 60.0%; Pred. No. 3.21e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 486 DGPGKPKQH 495  
:|||||:  
QY 1 DGPTGEPQOE 10

RESULT 9  
ID O44380 PRELIMINARY; PRT; 1398 AA.  
AC O44380;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE CIRCADIAN CLOCK PROTEIN.  
GN TIM.

OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
[1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 98033379.  
RA MYERS M.P., ROTHENFLUH A., CHANG M., YOUNG M.W.;  
RL "Comparison of chromosomal DNA composing timeless in Drosophila melanogaster and D. virilis suggests a new conserved structure for the TIMELESS protein.";  
RL NUCLEIC ACIDS RES. 25:4710-4714(1997).  
[2]

RP SEQUENCE OF 220-280 FROM N.A.  
RX MEDLINE; 96055118.  
RA MYERS M.P., WAGER-SMITH K., WESLEY C.S., YOUNG M.W., SEHGAL A.;  
RL "Positional cloning and sequence analysis of the Drosophila clock gene timeless.";  
RL SCIENCE 270:805-808(1995).  
[3]

RP SEQUENCE OF 220-280 FROM N.A.  
RA OUSLEY A., ZAFARULLAH K., CHEN Y., EMERSON M., HICKMAN L., SEHGAL A.;  
RL GENETICS 0:0-0(1998).  
DR EMBL; AF032401; G2655282; - -  
DR EMBL; AF032400; G2655282; JOINED.  
DR EMBL; AF038501; G2746731; - -  
SQ SEQUENCE 1398 AA; 156366 MW; 59A88D49 CRC32;

Query Match 70.1%; Score 47; DB 5; Length 1398;  
Best Local Similarity 60.0%; Pred. No. 3.21e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 519 DGPGKPKQH 528  
:|||||:  
QY 1 DGPTGEPQOE 10

RESULT 10  
ID O35966 PRELIMINARY; PRT; 231 AA.  
AC O35966;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE SEX DETERMINING PROTEIN.  
GN SRY.

OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
[1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 98043417.  
RC STRAIN-B6JEL-YBUB, CD-1 AND B6JEL-YKAM;  
RX MEDLINE; 98043417.

RA ALBRECHT K.H., EICHER E.M.;  
RL "DNA sequence analysis of Sry alleles (subgenus Mus) implicates Sry misregulation as the cause of C57BL/6J-Y(POS) sex reversal and defines the SRY functional unit.";  
RL GENETICS 147:1267-1277(1997).  
[2]

RP SEQUENCE FROM N.A.  
RC SPECIES-WESTERN EUROPEAN HOUSE MOUSE; STRAIN-B6-YFVB;

```

QY      1 DGTGPEQQE 10
| ||| |||

RESULT 12          PRELIMINARY; PRT;    233 AA.
ID   O35860;
AC   O35860;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DI   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SEX DETERMINING PROTEIN.
DN   SRY.
OS   MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
CC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]
RN
RP   SEQUENCE FROM N.A.
RC   STRAIN=WMP/EI;
RX   MEDLINE; 98043417.
RA   ALBRECHT K.H., EICHER E.M.;
RT   "DNA sequence analysis of Sry alleles (subgenus Mus) implicates
RT   misregulation as the cause of C57BL/6J-Y(POS) sex reversal and
RT   defines the SRY functional unit.";
RL   GENETICS 147:1267-1277(1997).
DR   ENBL; U70649; G2623363; -.
DP   PFAM; PF00505; HMG_box; 1.
SQ   SEQUENCE 233 AA; 28790 MW; 3F2451C1 CRC32;

Query Match      68.7%; Score 46; DB 11; Length 233;
Best Local Similarity 70.0%; Pred. No. 5.40e+00;
Matches       7; Conservative     1; Mismatches     2; Indels     0; Gaps     0

Db 137 DIPTGHPPQQ 146
| ||| |||

QY      1 DGTGPEQQE 10

RESULT 13          PRELIMINARY; PRT;    234 AA.
ID   O35859;
AC   O35859;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DI   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SEX DETERMINING PROTEIN.
DN   SRY.
OS   MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
CC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]
RN
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6JEI-YORB.
RX   MEDLINE; 98043417.
RA   ALBRECHT K.H., EICHER E.M.;
RT   "DNA sequence analysis of Sry alleles (subgenus Mus) implicates
RT   misregulation as the cause of C57BL/6J-Y(POS) sex reversal and
RT   defines the SRY functional unit.";
RL   GENETICS 147:1267-1277(1997).
DR   ENBL; U70646; G2623357; -.
DP   PFAM; PF00505; HMG_box; 1.
SQ   SEQUENCE 234 AA; 28918 MW; 69C12948 CRC32;

Query Match      68.7%; Score 46; DB 11; Length 234;
Best Local Similarity 70.0%; Pred. No. 5.40e+00;
Matches       7; Conservative     1; Mismatches     2; Indels     0; Gaps     0

Db 137 DIPTGHPPQQ 146
| ||| |||

QY      1 DGTGPEQQE 10

RESULT 14          PRELIMINARY; PRT;    234 AA.
ID   O35858;
AC   O35858;

```

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE SEX DETERMINING PROTEIN.  
 GN SRY.  
 OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B6JEL-YLIP;  
 RX MEDLINE: 98043417.  
 RA ALBRECHT K.H., EICHER E.M.;  
 RT "DNA sequence analysis of Sry alleles (subgenus Mus) implicates  
 RT misregulation as the cause of C57BL/6J-Y(POS) sex reversal and  
 RT defines the SRY functional unit.";  
 RL GENETICS 147:1267-1277(1997).  
 DR EMBL: U70645; G2623355; -.  
 DR PFAM: PF00505; HMG\_box; 1.  
 SQ SEQUENCE 234 AA; 28918 MW; 2446EF1B CRC32;  
  
 Query Match 58.7%; Score 46; DB 11; Length 234;  
 Best Local Similarity 70.0%; Pred. No. 5.40e+00;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
 Db 137 DIPTGHPQOO 146  
 QY 1 DGTGEPQOE 10  
  
 RESULT 15  
 ID O69004 PRELIMINARY; PRT; 238 AA.  
 AC O69004;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE FERRISIDOPHORE RECEPTOR-LIKE PROTEIN.  
 GN PIGC.  
 OS PSEUDOMONAS AERUGINOSA.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;  
 OC PSEUDOMONAS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PAOI;  
 RX MEDLINE: 96210657.  
 RA OCHSNER U.A., VASIL M.L.;  
 RT "Gene repression by the ferric uptake regulator in Pseudomonas  
 RT aeruginosa: cycle selection of iron-regulated genes.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:4409-4414(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PAOI;  
 RA OCHSNER U.A., JOHNSON Z., VASIL A.I., VASIL M.L.;  
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF060193; G3091148; -.  
 SQ SEQUENCE 238 AA; 25049 MW; 582EDF82 CRC32;  
  
 Query Match 68.7%; Score 46; DB 2; Length 238;  
 Best Local Similarity 55.6%; Pred. No. 5.40e+00;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
 Db 182 GSTGDPQPD 190  
 QY 2 GPTGEPQOE 10

Search completed: Thu Oct 21 15:43:42 1999  
 Job time : 48 secs.

\*\*\*\*\*

MSRCH\_PP

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 21 15:46:06 1999; MasPar time 3.97 Seconds  
Tubular output not generated. 53.512 Million cell updates/sec

Title: >US-09-040-485-9  
Description: (1-10) from US09040485.peg  
Perfect Score: 61  
Sequence: 1 QENPDSSEPV 10

Scoring table: PAM 150  
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseg35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 14.278; Variance 44.849; scale 0.318

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	43	70.5	216	33	C. felis esterase pro	2.19e+02
2	43	70.5	570	33	C. felis esterase, nF	2.19e+02
3	43	70.5	570	33	C. felis esterase pro	2.19e+02
4	43	70.5	570	33	C. felis esterase, nF	2.19e+02
5	43	70.5	570	33	C. felis esterase pro	2.19e+02
6	43	70.5	595	33	C. felis esterase, nF	2.19e+02
7	43	70.5	595	33	C. felis esterase, nF	2.19e+02
8	43	70.5	596	33	C. felis esterase, nF	2.19e+02
9	43	70.5	596	33	C. felis esterase, nF	2.19e+02
10	43	70.5	995	12	Agarase 0107 from Vib	2.19e+02
11	42	68.9	332	19	Beta-1-4-galactosyltr	2.79e+02
12	41	67.2	14	29	Cytotoxic Epstein-Bar	3.55e+02
13	41	67.2	84	13	N-terminal fragment	3.55e+02
14	41	67.2	154	13	N-terminal fragment	3.55e+02
15	41	67.2	188	13	N-terminal fragment	3.55e+02
16	41	67.2	228	13	N-terminal fragment	3.55e+02

17	41	67.2	289	13	R66796	Novel mouse proteogly	3.55e+02
18	41	67.2	311	39	W51198	Mouse syndecan-1.	3.55e+02
19	41	67.2	311	10	R55276	Syndecan protein.	3.55e+02
20	41	67.2	311	13	R66793	Novel mouse proteogly	3.55e+02
21	41	67.2	311	29	W47156	Mouse syndecan protei	3.55e+02
22	41	67.2	311	17	R87001	Mouse syndecan-1.	3.55e+02
23	41	67.2	625	33	W68294	Murine NF-KB receptor	3.55e+02
24	41	67.2	625	38	W83200	Murine osteoclast dif	3.55e+02
25	41	67.2	625	33	W69958	Murine NF-KB receptor	3.55e+02
26	41	67.2	919	33	W18580	Potato alpha-glucosid	3.55e+02
27	41	67.2	1461	34	W64468	Human secreted protei	3.55e+02
28	40	65.6	10	27	W42736	Antigenic decapeptide	4.50e+02
29	40	65.6	10	27	W42824	Antigenic decapeptide	4.50e+02
30	40	65.6	151	14	R79761	Human testis sperm zo	4.50e+02
31	40	65.6	151	27	W42642	Human sperm zona bind	4.50e+02
32	40	65.6	162	28	W42692	Recombinant human Spl	4.50e+02
33	40	65.6	163	28	W42693	Baboon sperm zona bin	4.50e+02
34	40	65.6	331	26	W29817	Mammalian AMPK-gamma	4.50e+02
35	40	65.6	461	27	W33890	Flea saliva protein p	4.50e+02
36	39	63.9	48	36	W79087	Human secreted protei	5.70e+02
37	39	63.9	145	32	W54376	IBDV VP5 protein.	5.70e+02
38	39	63.9	226	3	R13901	Nitrile hydratase L t	5.70e+02
39	39	63.9	260	21	W14422	Human DNase I variant	5.70e+02
40	39	63.9	496	22	W08605	Chimeric MoMLV and Ty	5.70e+02
41	39	63.9	866	19	W04185	Human interleukin-17	5.70e+02
42	39	63.9	866	33	W51272	Human interleukin-17	5.70e+02
43	39	63.9	1006	35	W72256	Human receptor type t	5.70e+02
44	39	63.9	2019	13	R67913	Cardiac sodium channe	5.70e+02
45	39	63.9	2020	2	R06584	Cardiac sodium channe	5.70e+02

ALIGNMENTS

RESULT 1

ID W57868 standard; Protein; 216 AA.  
AC W57868;  
DT 23-SEP-1998 (first entry)  
DE C. felis esterase protein.  
KW Esterase; flea; protective immune response; carboxylesterase; arthropod;  
KW haematophagous ectoparasite infestation.  
OS Ctenocephalides felis.  
PN W09821324-A1.  
PD 23-MAY-1998.  
PF 10-NOV-1997; U20598.  
PR 12-NOV-1996; US-747221.  
PI (HESK-) HESKA CORP.  
PI Brandt KS, Silver GM, Wisniewski N;  
DR WPI: 98-297929/26.  
DR N-PSDB: V40766.  
PT New nucleic acid encoding carboxyl-esterase(s) from fleas - useful,  
PT e.g. in vaccines, for preventing infestation by haematophagous  
PT ectoparasites, particularly on cats and dogs  
PS Claim 2; Page 204-205; 230pp; English.  
CC This sequence represents a flea esterase protein of the invention.  
CC When administered to animals, the protein induces a protective immune  
CC (antibody) response against carboxylesterase (CE), so they, or  
CC compositions containing CE proteins, are used, therapeutically or as  
CC vaccines, to protect particularly mammals and birds, specifically cats  
CC and dogs, against haematophagous ectoparasite infestation (HEP),  
CC specifically fleas. More generally the compositions can be used to treat  
CC arthropods generally, including pests of agricultural crops, trees, of  
CC stored goods etc., also those that are vectors of disease. Fragments of  
CC the DNA can be used as probes and primers for identification or  
CC production of nucleic acid. Antibodies against the protein can be used  
CC for passive immunisation; to screen expression libraries; to isolate the  
CC protein and to target cytotoxic compounds to HEP. The compounds  
CC containing CE are effective against both adult and larval stages; they  
CC target CEs, including juvenile hormone, that are involved in development,  
CC metamorphosis, feeding, digestion and reproduction.  
SQ Sequence 216 AA;

Query Match 70.5%; Score 43; DB 33; Length 216;  
Best Local Similarity 44.4%; Pred. No. 2.19e+02;

```

Matches      4;  Conservative      5;  Mismatches      0;  Indels      0;  Gaps      0;

Db 177 espesdpdi 185
    1:1:1:1:
Qy 2 ENPDSSEPV 10

RESULT 2
ID W57857 standard; Protein: 570 AA.
AC W57857;
DT 23-SEP-1998 (first entry)
DE C. felis esterase, nfe81710.
KW Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nfe81710.
OS Ctenocephalides felis.
PN W09821324-A1.
PD 22-MAY-1998.
PF 10-NOV-1997; U20598.
PR 12-NOV-1996; US-747221.
PA (HESK-) HESKA CORP.
PI Brandt KS, Silver GM, Wisniewski N;
DR WPI: 98-297929/26.
DR N-PSDB: V40746.
PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
PS Claim 1; Page 137-140; 230pp; English.
CC This sequence is the flea esterase protein, nfe81710, of the invention.
CC When administered to animals, the protein induces a protective immune
CC (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically cats
CC and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CES, including juvenile hormone, that are involved in development,
CC metamorphosis, feeding, digestion and reproduction.
SQ Sequence 570 AA:

Query Match 70.5%; Score 43; DB 33; Length 570;
Best Local Similarity 44.4%; Pred. No. 2.19e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 289 espesdpdi 297
    1:1:1:1:
Qy 2 ENPDSSEPV 10

RESULT 4
ID W57860 standard; Protein: 570 AA.
AC W57860;
DT 23-SEP-1998 (first entry)
DE C. felis esterase, nfe81710.
KW Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nfe81710.
OS Ctenocephalides felis.
PN W09821324-A1.
PD 22-MAY-1998.
PF 10-NOV-1997; U20598.
PR 12-NOV-1996; US-747221.
PA (HESK-) HESKA CORP.
PI Brandt KS, Silver GM, Wisniewski N;
DR WPI: 98-297929/26.
DR N-PSDB: V40751.
PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
PS Claim 1; Page 152-155; 230pp; English.
CC This sequence is the flea esterase protein, nfe81710, of the invention.
CC When administered to animals, the protein induces a protective immune
CC (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically cats
CC and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CES, including juvenile hormone, that are involved in development,
CC metamorphosis, feeding, digestion and reproduction.
SQ Sequence 570 AA:

Query Match 70.5%; Score 43; DB 33; Length 570;
Best Local Similarity 44.4%; Pred. No. 2.19e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 289 espesdpdi 297
    1:1:1:1:
Qy 2 ENPDSSEPV 10

RESULT 3
ID W57878 standard; Protein: 570 AA.
AC W57878;
DT 23-SEP-1998 (first entry)
DE C. felis esterase protein.
KW Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation.
OS Ctenocephalides felis.
PN W09821324-A1.
PD 22-MAY-1998.
PF 10-NOV-1997; U20598.
PR 12-NOV-1996; US-747221.
PA (HESK-) HESKA CORP.
PI Brandt KS, Silver GM, Wisniewski N;
DR WPI: 98-297929/26.
PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
PS Claim 2; Page 176-178; 230pp; English.

```



RESULT 5  
 ID W57879 standard; Protein; 570 AA.  
 AC W57879;  
 DT 23-SEP-1998 (first entry)  
 DE C. felis esterase protein.  
 KW Esterase; flea; protective immune response; carboxylesterase; arthropod;  
 KW haematophagous ectoparasite infestation.  
 OS Ctenocephalides felis.  
 PN W09821324-A1.  
 PD 22-MAY-1998.  
 PF 10-NOV-1997; U20598.  
 PR 12-NOV-1996; US-747221.  
 PA (HESK-) HESKA CORP.  
 PI Brandt KS, Silver GM, Wisniewski N;  
 DR WPI: 98-297929/26.  
 PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,  
 PT e.g. in vaccines, for preventing infestation by haematophagous  
 PT ectoparasites, particularly on cats and dogs  
 PS Claim 2; Page 178-180; 230pp; English.  
 CC This sequence represents a flea esterase protein of the invention.  
 CC When administered to animals, the protein induces a protective immune  
 CC (antibody) response against carboxylesterase (CE), so they, or  
 CC compositions containing CE proteins, are used, therapeutically or as  
 CC vaccines, to protect particularly mammals and birds, specifically cats  
 CC and dogs, against haematophagous ectoparasite infestation (HEP),  
 CC specifically fleas. More generally the compositions can be used to treat  
 CC arthropods generally, including pests of agricultural crops, trees,  
 CC stored goods etc., also those that are vectors of disease. Fragments of  
 CC the DNA can be used as probes and primers for identification or  
 CC for passive immunisation; to screen expression libraries; to isolate the  
 CC production of nucleic acid. Antibodies against the protein can be used  
 CC against haematophagous ectoparasite infestation (HEP).  
 CC protein and to target cytotoxic compounds to HEP. The compounds  
 CC containing CE are effective against both adult and larval stages; they  
 CC target CES, including juvenile hormone, that are involved in development,  
 CC metamorphosis, feeding, digestion and reproduction.  
 SQ Sequence 570 AA;  
 Query Match 70.5%; Score 43; DB 33; Length 570;  
 Best Local Similarity 44.4%; Pred. No. 2.19e+02;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 Db 289 espesdipi 297  
 QY 2 ENPDSSEPV 10  
 I:|:|:|:  
 RESULT 6  
 ID W57861 standard; Protein; 595 AA.  
 AC W57861;  
 DT 23-SEP-1998 (first entry)  
 DE C. felis esterase, nFE81785.  
 KW Esterase; flea; protective immune response; carboxylesterase; arthropod;  
 KW haematophagous ectoparasite infestation; nFE81785.  
 OS Ctenocephalides felis.  
 PN W09821324-A1.  
 PD 22-MAY-1998.  
 PF 10-NOV-1997; U20598.  
 PR 12-NOV-1996; US-747221.  
 PA (HESK-) HESKA CORP.  
 PI Brandt KS, Silver GM, Wisniewski N;  
 DR WPI: 98-297929/26.  
 DR N-PSDB: V40752.  
 PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,  
 PT e.g. in vaccines, for preventing infestation by haematophagous  
 PT ectoparasites, particularly on cats and dogs  
 PS Claim 1; Page 155-158; 230pp; English.  
 CC This sequence is the flea esterase protein, nFE81785, of the invention.  
 CC When administered to animals, the protein induces a protective immune  
 CC (antibody) response against carboxylesterase (CE), so they, or  
 CC compositions containing CE proteins, are used, therapeutically or as  
 CC vaccines, to protect particularly mammals and birds, specifically cats  
 CC and dogs, against haematophagous ectoparasite infestation (HEP),

CC specifically fleas. More generally the compositions can be used to treat  
 CC arthropods generally, including pests of agricultural crops, trees,  
 CC stored goods etc., also those that are vectors of disease. Fragments of  
 CC the DNA can be used as probes and primers for identification or  
 CC production of nucleic acid. Antibodies against the protein can be used  
 CC for passive immunisation; to screen expression libraries; to isolate the  
 CC protein and to target cytotoxic compounds to HEP. The compounds  
 CC containing CE are effective against both adult and larval stages; they  
 CC target CES, including juvenile hormone, that are involved in development,  
 CC metamorphosis, feeding, digestion and reproduction.  
 SQ Sequence 595 AA;  
 Query Match 70.5%; Score 43; DB 33; Length 595;  
 Best Local Similarity 44.4%; Pred. No. 2.19e+02;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 Db 314 espesdipi 322  
 QY 2 ENPDSSEPV 10  
 I:|:|:|:  
 RESULT 7  
 ID W57859 standard; Protein; 595 AA.  
 AC W57859;  
 DT 23-SEP-1998 (first entry)  
 DE C. felis esterase, nFE82801.  
 KW Esterase; flea; protective immune response; carboxylesterase; arthropod;  
 KW haematophagous ectoparasite infestation; nFE82801.  
 OS Ctenocephalides felis.  
 PN W09821324-A1.  
 PD 22-MAY-1998.  
 PF 10-NOV-1997; U20598.  
 PR 12-NOV-1996; US-747221.  
 PA (HESK-) HESKA CORP.  
 PI Brandt KS, Silver GM, Wisniewski N;  
 DR WPI: 98-297929/26.  
 DR N-PSDB: V40749.  
 PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,  
 PT e.g. in vaccines, for preventing infestation by haematophagous  
 PT ectoparasites, particularly on cats and dogs  
 PS Claim 2; Page 145-148; 230pp; English.  
 CC This sequence is the flea esterase protein, nFE82801, of the invention.  
 CC When administered to animals, the protein induces a protective immune  
 CC (antibody) response against carboxylesterase (CE), so they, or  
 CC compositions containing CE proteins, are used, therapeutically or as  
 CC vaccines, to protect particularly mammals and birds, specifically cats  
 CC and dogs, against haematophagous ectoparasite infestation (HEP),  
 CC specifically fleas. More generally the compositions can be used to treat  
 CC arthropods generally, including pests of agricultural crops, trees,  
 CC stored goods etc., also those that are vectors of disease. Fragments of  
 CC the DNA can be used as probes and primers for identification or  
 CC production of nucleic acid. Antibodies against the protein can be used  
 CC for passive immunisation; to screen expression libraries; to isolate the  
 CC protein and to target cytotoxic compounds to HEP. The compounds  
 CC containing CE are effective against both adult and larval stages; they  
 CC target CES, including juvenile hormone, that are involved in development,  
 CC metamorphosis, feeding, digestion and reproduction.  
 SQ Sequence 595 AA;  
 Query Match 70.5%; Score 43; DB 33; Length 595;  
 Best Local Similarity 44.4%; Pred. No. 2.19e+02;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 Db 314 espesdipi 322  
 QY 2 ENPDSSEPV 10  
 I:|:|:|:  
 RESULT 8  
 ID W57858 standard; Protein; 596 AA.  
 AC W57858;  
 DT 23-SEP-1998 (first entry)  
 DE C. felis esterase, nFE71788.

KW Esterase: flea; protective immune response; carboxylesterase; arthropod;  
 KW haematophagous ectoparasite infestation; nfe71788.  
 OS Ctenocephalides felis.  
 PN W09821324-A1.  
 PD 22-MAY-1998.  
 PF 10-NOV-1997; U20598.  
 PR 12-NOV-1996; US-747221.  
 PA (HESK-) HESKA CORP.  
 PI Brandt KS, Silver GM, Wisniewski N;  
 DR WPI; 98-297929/26.  
 DR N-PSDB; V40747.  
 PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,  
 PT e.g. in vaccines, for preventing infestation by haematophagous  
 PT ectoparasites, particularly on cats and dogs  
 PS Claim 1; Page 140-143; 230pp; English.  
 CC This sequence is the flea esterase protein, nfe71788, of the invention.  
 CC When administered to animals, the protein induces a protective immune  
 CC (antibody) response against carboxylesterase (CE), so they, or  
 CC compositions containing CE proteins, are used therapeutically or as  
 CC vaccines, to protect particularly mammals and birds, specifically cats  
 CC and dogs, against haematophagous ectoparasite infestation (HEP),  
 CC specifically fleas. More generally the compositions can be used to treat  
 CC arthropods generally, including pests of agricultural crops, trees,  
 CC stored goods etc., also those that are vectors of disease. Fragments of  
 CC the DNA can be used as probes and primers for identification or  
 CC production of nucleic acid. Antibodies against the protein can be used  
 CC for passive immunisation; to screen expression libraries; to isolate the  
 CC protein and to target cytotoxic compounds to HEP. The compounds  
 CC containing CE are effective against both adult and larval stages; they  
 CC target CEs, including juvenile hormone, that are involved in development,  
 CC metamorphosis, feeding, digestion and reproduction.  
 SQ Sequence 596 AA;

Query Match 70.5%; Score 43; DB 33; Length 596;  
 Best Local Similarity 44.4%; Pred. NO. 2.19e+02;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 315 espesdipi 323  
 |::|::|  
 QY 2 ENPDSSEPV 10

RESULT 9  
 ID W57856 standard; Protein; 596 AA.  
 AC W57856;  
 DT 23-SEP-1998 (first entry)  
 DE C. felis esterase, nfe72836.  
 KW Esterase: flea; protective immune response; carboxylesterase; arthropod;  
 KW haematophagous ectoparasite infestation; nfe72836.  
 OS Ctenocephalides felis.  
 PN W09821324-A1.  
 PD 22-MAY-1998.  
 PF 10-NOV-1997; U20598.  
 PR 12-NOV-1996; US-747221.  
 PA (HESK-) HESKA CORP.  
 PI Brandt KS, Silver GM, Wisniewski N;  
 DR WPI; 98-297929/26.  
 DR N-PSDB; V40744.  
 PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,  
 PT e.g. in vaccines, for preventing infestation by haematophagous  
 PT ectoparasites, particularly on cats and dogs  
 PS Claim 2; Page 129-133; 230pp; English.  
 CC This sequence is the flea esterase protein, nfe72836, of the invention.  
 CC When administered to animals, the protein induces a protective immune  
 CC (antibody) response against carboxylesterase (CE), so they, or  
 CC compositions containing CE proteins, are used therapeutically or as  
 CC vaccines, to protect particularly mammals and birds, specifically cats  
 CC and dogs, against haematophagous ectoparasite infestation (HEP),  
 CC specifically fleas. More generally the compositions can be used to treat  
 CC arthropods generally, including pests of agricultural crops, trees,  
 CC stored goods etc., also those that are vectors of disease. Fragments of  
 CC the DNA can be used as probes and primers for identification or  
 CC production of nucleic acid. Antibodies against the protein can be used

CC for passive immunisation; to screen expression libraries; to isolate the  
 CC protein and to target cytotoxic compounds to HEP. The compounds  
 CC containing CE are effective against both adult and larval stages; they  
 CC target CEs, including juvenile hormone, that are involved in development,  
 CC metamorphosis, feeding, digestion and reproduction.  
 SQ Sequence 596 AA;

Query Match 70.5%; Score 43; DB 33; Length 596;  
 Best Local Similarity 44.4%; Pred. NO. 2.19e+02;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 315 espesdipi 323  
 |::|::|  
 QY 2 ENPDSSEPV 10

RESULT 10  
 ID R60812 standard; Protein; 995 AA.  
 AC R60812;  
 DT 20-JUL-1995 (first entry)  
 DE Agarase 0107 from Vibrio species.  
 KW Agarase; Vibrio; PCR; amplification; amplify; primer; probe; E.coli;  
 KW cation-exchange; chromatograph; neogaroooligosaccharide.  
 OS Vibrio sp. Jn0107.  
 FH Key Location/Qualifiers  
 FT peptide 1..20  
 FT /label= signal peptide  
 FT protein 21..995  
 FT /label= mature peptide  
 PN J06284888-A.  
 PD 11-OCT-1994.  
 PF 01-APR-1993; 096549.  
 PR 01-APR-1993; JP-096549.  
 PA (NISB) JAPAN TOBACCO INC.  
 DR WPI; 94-362595/45.  
 DR N-PSDB; 073755.  
 PT DNA encoding protein with agarase activity - useful for the  
 PT industrial production of a large amount of  
 PT neo-agaro-oligosaccharide  
 PS Claim 1; Page 8-12; 13pp; Japanese.  
 CC The sequence of a novel protein containing agarase activity from  
 CC Vibrio species. The gene of 3205 bp encodes a protein of 995 amino  
 CC acids. The gene was obtained by PCR amplification from a Vibrio genomic  
 CC library using primers and probes based on the amino acid sequence  
 CC (R60813-21) from protein sequencing of partially degraded agarase 0107.  
 CC The gene was then introduced into E.coli JM83. The expressed protein  
 CC could then be recovered by cation-exchange chromatography. The protein  
 CC can be used in the large scale manufacture of neogaroooligosaccharide.  
 SQ Sequence 995 AA;

Query Match 70.5%; Score 43; DB 12; Length 995;  
 Best Local Similarity 66.7%; Pred. NO. 2.19e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 947 eapdpgepv 955  
 |::|::|  
 QY 2 ENPDSSEPV 10

RESULT 11  
 ID W06491 standard; Protein; 332 AA.  
 AC W06491;  
 DT 05-FEB-1997 (first entry)  
 DE Beta-1-4-galactosyltransferase-related protein #2.  
 KW Murine; beta-1,4-galactosyltransferase-related protein; sterility;  
 KW fertilisation; F9 cancer cell; Huynh's method.  
 OS Mus musculus.  
 PN J08196279-A.  
 PD 06-AUG-1996.  
 PF 25-JAN-1995; 009642.  
 PR 25-JAN-1995; JP-009642.  
 PA (MITK) MITSUI TOATSU CHEM INC.  
 PA (MURA) MURAMATSU T.

DR WPI: 96-406013/41.  
 DR N-PSDB: T45082.  
 PT DNA sequence encoding beta-1,4-galactosyltransferase-related  
 PT protein - useful for sterility diagnosis, and for assisting or  
 PT inhibiting fertilisation  
 PS Claim 4; Page 7-9; lipp; Japanese.  
 CC The sequences given in W06490-91 represent two clones of murine  
 CC beta-1,4-galactosyltransferase-related proteins. These proteins  
 CC can be used as diagnostic agents for various diseases. They are  
 CC esp. useful in the diagnosis of sterility and in the aiding and  
 CC inhibiting of fertilisation. The cDNA's encoding the two beta-1,4-  
 CC galactosyltransferase-related proteins were isolated from F9 cancer  
 CC cells according to Huynh's method.  
 SQ Sequence 332 AA;

Query Match 58.9%; Score 42; DB 19; Length 332;  
 Best Local Similarity 55.6%; Pred. No. 2.79e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 226 dspdsdvpv 234  
 QY :|||:|

RESULT 12  
 ID W40864 standard; peptide; 14 AA.  
 AC W40864;  
 DT 09-JUN-1998 (first entry)  
 DE Cytotoxic Epstein-Barr T-cell epitope 37.  
 KW Herpes simplex virus type 4; Epstein-Barr virus; EBV; cytotoxic;  
 KW T-cell epitope; nuclear antigen; human leukocyte antigen; HLA; vaccine;  
 KW tetanus toxoid; diphtheria toxoid; Bordetella pertussis;  
 KW poliovirus antigen.  
 OS Herpes simplex virus type 4.  
 PN W09745444-A1.  
 PD 04-DEC-1997.  
 PF 23-MAY-1997; AU0328.  
 PR 24-MAY-1996; AU-000073.  
 PA (BIOT-) BIOTECH AUSTRALIA PTY LTD.  
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.  
 PA (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.  
 PA (CSLC-) CSL LTD.  
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.  
 PA (UYNE) UNIV MELBOURNE.  
 PI Burrows SR, Kerr BM, Khanna R, Misko IS, Moss DJ;  
 DR WPI: 98-032576/03.  
 PT Cytotoxic Epstein-Barr virus T cell epitope - useful to prepare  
 PT vaccine to prevent and treat infection  
 PS Claim 3; Page 32; 4lpp; English.  
 CC This amino acid sequence is a cytotoxic Epstein-Barr virus (EBV) T-cell  
 CC epitope of the EBV nuclear antigen 3A (EBNA3A), which binds the human  
 CC leukocyte antigen acceptor A11 (HLA A11). It is used to prepare a  
 CC vaccine which may include the cytotoxic EBV T-cell epitope (W40828-  
 CC W40846), or a nucleic acid sequence encoding it. The vaccine produced  
 CC comprises or encodes at least one antigen (W40847-W40876) to which the  
 CC individual will mount an anamnestic response, e.g. a tetanus toxoid,  
 CC diphtheria toxoid, Bordetella pertussis antigen, poliovirus antigen,  
 CC purified protein derivative, glycoprotein 350 protein, and/or helper  
 CC epitope.  
 SQ Sequence 14 AA;

Query Match 67.2%; Score 41; DB 29; Length 14;  
 Best Local Similarity 55.6%; Pred. No. 3.55e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1 depastepv 9  
 QY :|||:|

RESULT 13  
 ID R66800 standard; Peptide; 84 AA.  
 AC R66800;

DT 11-SEP-1995 (first entry)  
 DE N-terminal fragment (a.a. 1-84) of mature mouse syndecan-1.  
 KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;  
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;  
 KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;  
 KW receptor; drug; antibody; diagnostic agent.  
 OS Mus musculus.  
 PN W09500633-A.  
 PD 05-JAN-1995.  
 PF 17-JUN-1994; U06920.  
 PR 17-JUN-1993; US-078683.  
 PA (CHIL-) CHILDRENS MEDICAL CENT.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Bernfield M, Kato M, Saunders S;  
 DR WPI: 95-052071/07.  
 PT DNA and protein sequences for recombinant syndecan-derived  
 PT proteoglycans - comprising a core protein having glycosylation  
 PT sites for heparin sulphate glycosaminoglycan side chains.  
 PS Claim 15; Page 82; 97pp; English.  
 CC Peptides R66797-802 correspond to varying length fragments of the  
 CC N-terminal sequence of the mature mouse syndecan-1 (see R66793 and  
 CC R66796). The sequence given here covers the amino acids from 1-84 of the  
 CC mature protein sequence. The fragments are portions of the extracellular  
 CC domain of the protein and contain a heparan sulphate glycosaminoglycan  
 CC attachment site (R66795). The functional domains, esp. the soluble  
 CC extracellular or heparan binding site, of the syndecan molecules (see  
 CC R66797-812 and R66818) can be used to construct chimaeras by linking them  
 CC to biological effector molecules, cell surface receptors, drugs,  
 CC antibodies, diagnostic agents or components of microorganisms.  
 SQ Sequence 84 AA;

Query Match 67.2%; Score 41; DB 13; Length 84;  
 Best Local Similarity 55.6%; Pred. No. 3.55e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpegepv 82  
 QY |::|:|

RESULT 14  
 ID R66801 standard; Peptide; 154 AA.  
 AC R66801;  
 DT 11-SEP-1995 (first entry)  
 DE N-terminal fragment (a.a. 1-154) of mature mouse syndecan-1.  
 KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;  
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;  
 KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;  
 KW receptor; drug; antibody; diagnostic agent.  
 OS Mus musculus.  
 PN W09500633-A.  
 PD 05-JAN-1995.  
 PF 17-JUN-1994; U06920.  
 PR 17-JUN-1993; US-078683.  
 PA (CHIL-) CHILDRENS MEDICAL CENT.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Bernfield M, Kato M, Saunders S;  
 DR WPI: 95-052071/07.  
 PT DNA and protein sequences for recombinant syndecan-derived  
 PT proteoglycans - comprising a core protein having glycosylation  
 PT sites for heparin sulphate glycosaminoglycan side chains.  
 PS Claim 15; Page 82; 97pp; English.  
 CC Peptides R66797-802 correspond to varying length fragments of the  
 CC N-terminal sequence of the mature mouse syndecan-1 (see R66793 and  
 CC R66796). The sequence given here covers the amino acids from 1-154 of the  
 CC mature protein sequence. The fragments are portions of the extracellular  
 CC domain of the protein and contain a heparan sulphate glycosaminoglycan  
 CC attachment site (R66795). The functional domains, esp. the soluble  
 CC extracellular or heparan binding site, of the syndecan molecules (see  
 CC R66797-812 and R66818) can be used to construct chimaeras by linking them  
 CC to biological effector molecules, cell surface receptors, drugs,  
 CC antibodies, diagnostic agents or components of microorganisms.  
 SQ Sequence 154 AA;

Query Match 67.2%; Score 41; DB 13; Length 154;  
 Best Local Similarity 55.6%; Pred. NO. 3.55e-02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeesepv 82  
 I:I: :|||  
 QY 2 ENPDSSEPV 10

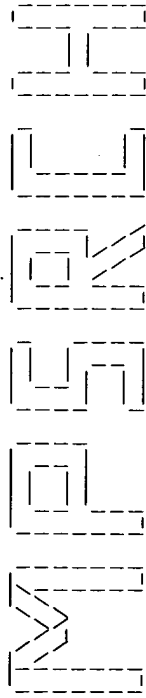
RESULT 15  
 ID R66802 standard; Peptide; 188 AA.  
 AC R66802;  
 DT 11-SEP-1995 (first entry)  
 DE N-terminal fragment (a.a. 1-188) of mature mouse syndecan-1.  
 KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;  
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;  
 KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;  
 KW receptor; drug; antibody; diagnostic agent.  
 OS Mus musculus.  
 PN W09500633-A.  
 PD 05-JAN-1995.  
 PF 17-JUN-1994; U06920.  
 PR 17-JUN-1993; US-078683.  
 PA (CHIL-) CHILDRENS MEDICAL CENT.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Bernfield M, Kato M, Saunders S;  
 DR WPI; 95-052071/07.  
 PT DNA and protein sequences for recombinant syndecan-derived  
 PT proteoglycans - comprising a core protein having glycosylation  
 PT sites for heparin sulphate glycosaminoglycan side chains.  
 PS Claim 15; Page 82; 97pp; English.  
 CC Peptides R66797-802 correspond to varying length fragments of the  
 CC N-terminal sequence of the mature mouse syndecan-1 (see R66793 and  
 CC R66796). The sequence given here covers the amino acids from 1-188 of the  
 CC mature protein sequence. The fragments are portions of the extracellular  
 CC domain of the protein and contain a heparan sulphate glycosaminoglycan  
 CC attachment site (R66795). The functional domains, esp. the soluble  
 CC extracellular or heparan binding site, of the syndecan molecules (see  
 CC R66797-812 and R66818) can be used to construct chimaeras by linking them  
 CC to biological effector molecules, cell surface receptors, drugs,  
 CC antibodies, diagnostic agents or components of microorganisms.  
 SQ Sequence 188 AA;

Query Match 67.2%; Score 41; DB 13; Length 188;  
 Best Local Similarity 55.6%; Pred. NO. 3.55e-02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeesepv 82  
 I:I: :|||  
 QY 2 ENPDSSEPV 10

Search completed: Thu Oct 21 15:46:35 1999  
 Job time : 29 secs.

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 21 15:49:10 1999; MasPar time 1.51 Seconds  
Tabular output not generated. 77.745 Million cell updates/sec

Title: >US-09-040-485-9  
Description: (1-10) from US09040485.pep  
Perfect Score: 61  
Sequence: 1 QENPDSEPV 10

Scoring table: PAM 150  
Gap 15

Searched: 119857 seqs, 11713122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 13.481; Variance 42.498; scale 0.317

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.
1	41	67.2	311	1 US-08-078- Sequence 2, Applicatio	1.79e+02
2	41	67.2	311	2 US-08-489- Sequence 6, Applicatio	1.79e+02
3	41	67.2	311	1 US-08-472- Sequence 2, Applicatio	1.79e+02
4	41	67.2	313	1 US-08-078- Sequence 4, Applicatio	1.79e+02
5	40	65.6	151	2 US-08-436- Sequence 2, Applicatio	2.28e+02
6	40	65.6	151	1 US-08-166- Sequence 2, Applicatio	2.28e+02
7	40	65.6	151	2 US-08-436- Sequence 2, Applicatio	2.28e+02
8	40	65.6	331	2 US-08-878- Sequence 21, Applicati	2.28e+02
9	40	65.6	415	2 US-08-381- Sequence 2, Applicatio	2.28e+02
10	40	65.6	461	2 US-08-630- Sequence 68, Applicati	2.28e+02
11	40	65.6	461	2 US-09-005- Sequence 68, Applicati	2.28e+02
12	39	63.9	145	2 US-08-708- Sequence 32, Applicati	2.91e+02
13	39	63.9	203	2 US-08-598- Sequence 4, Applicatio	2.91e+02
14	39	63.9	226	1 US-08-028- Sequence 4, Applicatio	2.91e+02
15	39	63.9	226	2 US-08-461- Sequence 4, Applicatio	2.91e+02
16	39	63.9	866	2 US-08-620- Sequence 10, Applicati	2.91e+02
17	38	62.3	147	2 US-08-598- Sequence 2, Applicatio	3.70e+02
18	38	62.3	153	2 US-08-606- Sequence 27, Applicati	3.70e+02
19	38	62.3	591	2 US-08-889- Sequence 1, Applicatio	3.70e+02
20	38	62.3	605	2 US-08-889- Sequence 2, Applicatio	3.70e+02
21	38	62.3	1086	4 5386025-8 Patent No. 5386025	3.70e+02
22	38	62.3	1103	2 US-08-223- Sequence 53, Applicati	3.70e+02
23	38	62.3	1103	2 US-08-455- Sequence 53, Applicati	3.70e+02

ALIGNMENTS

RESULT	1				
ID	US-08-078-683A-2	STANDARD;	PRT;	311 AA.	
XX	xxxxxx				
AC					
XX					
DT					
XX					
DE	Sequence 2, Application US/08078683A				
XX					
CC	Sequence 2, Application US/08078683A				
CC	Patent No. 5486599				
CC	GENERAL INFORMATION:				
CC	APPLICANT: Saunders, Scott				
CC	APPLICANT: Bernfield, Merton				
CC	APPLICANT: Kato, Masato				
CC	TITLE OF INVENTION: Construction and Use of Synthetic				
CC	TITLE OF INVENTION: Constructs Encoding Syndecan				
CC	NUMBER OF SEQUENCES: 43				
CC	CORRESPONDENCE ADDRESS:				
CC	ADDRESSEE: LAHIVE & COCKFIELD				
CC	STREET: 60 State Street				
CC	CITY: Boston				
CC	STATE: MA				
CC	COUNTRY: USA				
CC	ZIP: 02109				
CC	COMPUTER READABLE FORM:				
CC	MEDIUM TYPE: Floppy disk				
CC	COMPUTER: IBM PC compatible				
CC	OPERATING SYSTEM: PC-DOS/MS-DOS				
CC	SOFTWARE: ASCII (text)				
CC	CURRENT APPLICATION DATA:				
CC	APPLICATION NUMBER: US/08/078,683A				
CC	FILING DATE: 17-JUN-1993				
CC	CLASSIFICATION: 435				
CC	ATTORNEY/AGENT INFORMATION:				
CC	NAME: Vincent, Matthew P.				
CC	REGISTRATION NUMBER: 36,709				
CC	REFERENCE/DOCKET NUMBER: CME-062				
CC	TELECOMMUNICATION INFORMATION:				
CC	TELEPHONE: (617) 227-7400				
CC	TELEFAX: (617) 227-5941				
CC	INFORMATION FOR SEQ ID NO: 2:				
CC	SEQUENCE CHARACTERISTICS:				
CC	LENGTH: 311 amino acids				
CC	TYPE: amino acid				
CC	TOPOLOGY: linear				
CC	MOLECULE TYPE: protein				

SQ SEQUENCE 311 AA; 32904 MW; 499722 CN;

Query Match 67.2%; Score 41; DB 1; Length 311;  
Best Local Similarity 55.8%; Pred. No. 1.79e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 EKPEGEVP 104  
I:|: :|||  
QY 2 ENPDSSEPV 10

RESULT 2  
ID US-08-488-199-6 STANDARD; PRT; 311 AA.  
XX AC xxxxxx  
XX DT  
XX DE Sequence 2, Application US/08472217  
XX DE Sequence 2, Application US/08472217  
XX DE Patent No. 5726058  
XX CC GENERAL INFORMATION:  
XX CC APPLICANT: Alanen-Kurki, Leena  
XX CC APPLICANT: Auvinen, Petri  
XX CC APPLICANT: Jaakkola, Panu  
XX CC APPLICANT: Jalkanen, Markku  
XX CC APPLICANT: Lepp, Sirpa  
XX CC APPLICANT: Mali, Markku  
XX CC APPLICANT: Vihinen, Tapani  
XX CC APPLICANT: W rri, Anni  
XX CC TITLE OF INVENTION: Syndecan Stimulation Of Cellular  
XX CC TITLE OF INVENTION: Differentiation  
XX CC NUMBER OF SEQUENCES: 4  
XX CC CORRESPONDENCE ADDRESS:  
XX CC ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
XX CC STREET: 1100 New York Avenue, Suite 600  
XX CC CITY: Washington  
XX CC STATE: D.C.  
XX CC COUNTRY: U.S.A.  
XX CC ZIP: 20005  
XX CC COMPUTER READABLE FORM:  
XX CC MEDIUM TYPE: Floppy disk  
XX CC COMPUTER: IBM PC compatible  
XX CC OPERATING SYSTEM: PC-DOS/MS-DOS  
XX CC SOFTWARE: Patent In Release #1.0, Version #1.25  
XX CC CURRENT APPLICATION DATA:  
XX CC APPLICATION NUMBER: US/08/472,217  
XX CC FILING DATE: 07-JUN-1995  
XX CC CLASSIFICATION: 514  
XX CC PRIOR APPLICATION DATA:  
XX CC APPLICATION NUMBER: US 08/206,186  
XX CC FILING DATE: 07-MAR-1994  
XX CC PRIOR APPLICATION DATA:  
XX CC APPLICATION NUMBER: US 07/988,427  
XX CC FILING DATE: 01-DEC-1992  
XX CC ATTORNEY/AGENT INFORMATION:  
XX CC NAME: Cimbala, Michele A.  
XX CC REGISTRATION NUMBER: 33,851  
XX CC REFERENCE/DOCKET NUMBER: 1102.0050003  
XX CC TELECOMMUNICATION INFORMATION:  
XX CC TELEPHONE: (202) 371-2600  
XX CC TELEFAX: (202) 371-2540  
XX CC INFORMATION FOR SEQ ID NO: 2:  
XX CC SEQUENCE CHARACTERISTICS:  
XX CC LENGTH: 311 amino acids  
XX CC TYPE: amino acid  
XX CC TOPOLOGY: linear  
XX CC MOLECULE TYPE: protein  
SQ SEQUENCE 311 AA; 32904 MW; 499722 CN;

Query Match 67.2%; Score 41; DB 1; Length 311;  
Best Local Similarity 55.8%; Pred. No. 1.79e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 EKPEGEVP 104  
I:|: :|||  
QY 2 ENPDSSEPV 10

RESULT 4  
ID US-08-078-683A-4 STANDARD; PRT; 313 AA.  
XX AC xxxxxx  
XX DT

SQ SEQUENCE 311 AA; 32904 MW; 499722 CN;

Query Match 67.2%; Score 41; DB 2; Length 311;  
Best Local Similarity 55.6%; Pred. No. 1.79e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 EKPEGEVP 104  
I:|: :|||  
QY 2 ENPDSSEPV 10

RESULT 3  
ID US-08-472-217-2 STANDARD; PRT; 311 AA.

```
XX Sequence 4, Application US/08078683A
DE
XX
CC Sequence 4, Application US/08078683A
CC Patent No. 5486599
CC GENERAL INFORMATION:
CC APPLICANT: Saunders, Scott
CC APPLICANT: Bernfield, Merton
CC APPLICANT: Kato, Masato
CC TITLE OF INVENTION: Construction and Use of Synthetic
CC TITLE OF INVENTION: Constructs Encoding Syndecan
CC NUMBER OF SEQUENCES: 43
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 State Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII (text)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/078,683A
CC FILING DATE: 17-JUN-1993
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Vincent, Matthew P.
CC REGISTRATION NUMBER: 36,709
CC REFERENCE/DOCKET NUMBER: CME-062
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 313 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 313 AA; 33213 MW; 500523 CN;

Query Match 67.2%; Score 41; DB 1; Length 313;
Best Local Similarity 55.6%; Pred. No. 1.79e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 97 EKPEGEVP 105
QY 2 ENPDSEVP 10

RESULT 5
ID US-08-436-883B-2 STANDARD; PRT; 151 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
XX Sequence 2, Application US/08436883B
XX Patent No. 5820861
XX GENERAL INFORMATION:
XX APPLICANT: O'Rand, Michael G.
XX APPLICANT: Widgren, Esther E.
XX APPLICANT: Richardson, Richard T.
XX APPLICANT: Lea, Isabel
XX TITLE OF INVENTION: Sperm Antigen Corresponding to a Sperm
XX TITLE OF INVENTION: Zona Binding Protein Autoantigenic Epitope
XX NUMBER OF SEQUENCES: 60
XX CORRESPONDENCE ADDRESS:

CC Sequence 4, Application US/08078683A
CC Patent No. 5486599
CC GENERAL INFORMATION:
CC APPLICANT: Saunders, Scott
CC APPLICANT: Bernfield, Merton
CC APPLICANT: Kato, Masato
CC TITLE OF INVENTION: Construction and Use of Synthetic
CC TITLE OF INVENTION: Constructs Encoding Syndecan
CC NUMBER OF SEQUENCES: 43
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 State Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII (text)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/078,683A
CC FILING DATE: 17-JUN-1993
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Vincent, Matthew P.
CC REGISTRATION NUMBER: 36,709
CC REFERENCE/DOCKET NUMBER: CME-062
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 313 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 313 AA; 33213 MW; 500523 CN;

Query Match 65.6%; Score 40; DB 2; Length 151;
Best Local Similarity 55.6%; Pred. No. 2.28e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 77 QEPPEKSDP 85
QY 1 QENPDSEVP 9

RESULT 6
ID US-08-166-195A-2 STANDARD; PRT; 151 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
XX Sequence 2, Application US/08166195A
XX Patent No. 5480799
XX GENERAL INFORMATION:
XX APPLICANT: O'Rand, Michael G.
XX APPLICANT: Widgren, Esther E.
XX APPLICANT: Richardson, Richard T.
XX APPLICANT: Lea, Isabel
XX TITLE OF INVENTION: Sperm Antigen Corresponding to a
XX TITLE OF INVENTION: Sperm Zona Binding Protein Autoantigenic Epitope
XX NUMBER OF SEQUENCES: 51
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Kenneth D. Sibley
XX STREET: P.O. Box 34009
XX CITY: Charlotte
XX STATE: No. 5480799th Carolina
XX COUNTRY: USA
XX ZIP: 28234
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX COMPUTER: IBM PC compatible
XX OPERATING SYSTEM: PC-DOS/MS-DOS
XX SOFTWARE: PatentIn Release #1.0, Version #1.25
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: US/08/166,195A
XX FILING DATE: 10 DEC 1993
```

```

SQ SEQUENCE 151 AA; 17406 MW; 104641 CN;

Query Match 65.6%; Score 40; DB 2; Length 151;
Best Local Similarity 55.6%; Pred. No. 2.28e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 77 QEPPEKSDP 85
   ||| :||
QY 1 QENPDSSEP 9

RESULT 8
ID US-08-878-989-21 STANDARD; PRT; 331 AA.
XX AC
AC xxxxxx
DT XX
DT XX
DE Sequence 21, Application US/08878989
XX CC
CC Sequence 21, Application US/08878989
CC Patent No. 5885803
CC GENERAL INFORMATION:
CC APPLICANT: Bandman, Olga
CC APPLICANT: Hillman, Jennifer L.
CC APPLICANT: Corley, Neil C.
CC APPLICANT: Guegler, Karl G.
CC APPLICANT: Lal, Preeti
CC APPLICANT: Goli, Surya K.
CC APPLICANT: Shah, Purvi
CC TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
CC TITLE OF INVENTION: KINASES
CC NUMBER OF SEQUENCES: 21
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.
CC STREET: 3174 Porter Drive
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94304
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/878,989
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Billings, Lucy J J
CC REGISTRATION NUMBER: 36,749
CC REFERENCE/DOCKET NUMBER: PF-0321 US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-855-0555
CC TELEFAX: 415-845-4166
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 21:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 331 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC LIBRARY: GenBank
CC CLONE: 1335856
CC SEQUENCE 331 AA; 37579 MW; 613891 CN;

Query Match 65.6%; Score 40; DB 2; Length 331;
Best Local Similarity 50.0%; Pred. No. 2.28e+02;

```



Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 19 QETPESNSV 28  
|||:|:|:|  
QY 1 QENPDSSEPV 10

RESULT 9  
ID US-08-381-936-2 STANDARD; PRT; 415 AA.  
XX AC xxxxxx

Sequence 2, Application US/08381936

Sequence 2, Application US/08381936

Patent No. 5792923

GENERAL INFORMATION:

APPLICANT: ROBER, Manuela

APPLICANT: GEIER, Gebhardt

APPLICANT: GEIDER, Klaus

APPLICANT: WILLMITZER, Iothar

TITLE OF INVENTION: DNA sequences which lead to the

TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing

TITLE OF INVENTION: these sequences as well as a process for preparing

TITLE OF INVENTION: transgenic plants.

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ostrolenko, Faber, Gerb & Soffen

STREET: 1180 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10036-8403

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/381,936

FILING DATE: 09-FEB-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 93 02110

FILING DATE: 09-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE P42270618

FILING DATE: 08-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Mellman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/951-108

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888

TELEX: 236925

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 415 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 415 AA; 46408 MW; 961134 CN;

Query Match 65.6%; Score 40; DB 2; Length 415;

Best Local Similarity 30.0%; Pred. No. 2.28e+02;

Matches 3; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 21 EDDPTTTPV 30

|||:|:|:|

QY 1 QENPDSSEPV 10

RESULT 10  
ID US-08-630-822A-68 STANDARD; PRT; 461 AA.  
XX AC xxxxxx

Sequence 68, Application US/08630822A

Sequence 68, Application US/08630822A

Patent No. 5840695

GENERAL INFORMATION:

APPLICANT: FRANK, GLENN R.

APPLICANT: HUNTER, SHIRLEY WU

APPLICANT: WALLENFELS, LYNDIA

TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS

TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross P.C.

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: U.S.A.

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/630,822A

FILING DATE: 11-APR-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CONNELL, GARY J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-17-C3

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:

LENGTH: 461 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 461 AA; 52143 MW; 1160977 CN;

Query Match 65.6%; Score 40; DB 2; Length 461;

Best Local Similarity 55.6%; Pred. No. 2.28e+02;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 344 DEKPEPSEP 352

|||:|:|:|

QY 1 QENPDSSEP 9

RESULT 11  
ID US-09-005-069-68 . STANDARD; PRT; 461 AA.  
XX AC xxxxxx

Sequence 68, Application US/09005069

Sequence 68, Application US/09005069

Patent No. 5932470

GENERAL INFORMATION:

APPLICANT: FRANK, GLENN R.



CC TELEX: 66141 PENNIE  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 203 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 203 AA; 22922 MW; 208307 CN;

Query Match 63.9%; Score 39; DB 2; Length 203;  
Best Local Similarity 30.0%; Pred. No. 2.91e+02;  
Matches 3; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 89 QDGPAGQSV 98

QY 1 QENPDSSEPV 10

RESULT 14  
ID US-08-028-463-4 STANDARD; PRT: 226 AA.

XX xxxxxx

Sequence 4, Application US/08028463

Sequence 4, Application US/08028463

Patent No. 5731176

GENERAL INFORMATION:

APPLICANT: BEPPU, TERUHIKO

APPLICANT: YAMADA, HIDEAKI

APPLICANT: NAGASAWA, TORU

APPLICANT: HORINOUCHI, SUCHAYU

APPLICANT: NISHIYAMA, MAKATO

TITLE OF INVENTION: DNA FRAGMENT ENCODING A POLYPEPTIDE

TITLE OF INVENTION: HAVING NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING

TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE

TITLE OF INVENTION: TRANSFORMANT

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 AVENUE OF THE AMERICAS

CITY: NEW YORK

STATE: N.Y.

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/028,463

FILING DATE: 09-MAR-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/694,747

FILING DATE: 02-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: MISROCK, S. LESLIE

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7005-023-999

TELEPHONE: 212 790-9090

TELEFAX: 212 869-9741

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 226 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

CC ORIGINAL SOURCE:  
CC ORGANISM: Rhodococcus rhodochrous  
CC STRAIN: J-1 (FERM BP-1478)  
SQ SEQUENCE 226 AA; 25201 MW; 263723 CN;

Query Match 63.9%; Score 39; DB 1; Length 226;  
Best Local Similarity 62.5%; Pred. No. 2.91e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 17 KPESDEPV 24

QY 3 NPDSEPV 10

RESULT 15

ID US-08-461-836-4 STANDARD; PRT: 226 AA.

XX xxxxxx

Sequence 4, Application US/08461836

Sequence 4, Application US/08461836

Patent No. 5753472

GENERAL INFORMATION:

APPLICANT: BEPPU, TERUHIKO

APPLICANT: YAMADA, HIDEAKI

APPLICANT: NAGASAWA, TORU

APPLICANT: HORINOUCHI, SUCHAYU

APPLICANT: NISHIYAMA, MAKATO

TITLE OF INVENTION: DNA FRAGMENT ENCODING A POLYPEPTIDE

TITLE OF INVENTION: HAVING NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING

TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING T

TITLE OF INVENTION: TRANSFORMANT

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 AVENUE OF THE AMERICAS

CITY: NEW YORK

STATE: N.Y.

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/461,836

FILING DATE: 05-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/028,463

FILING DATE: 09-MAR-1993

APPLICATION NUMBER: US 07/694,747

FILING DATE: 02-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: MISROCK, S. LESLIE

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7005-023-999

TELEPHONE: 212 790-9090

TELEFAX: 212 869-9741

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 226 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

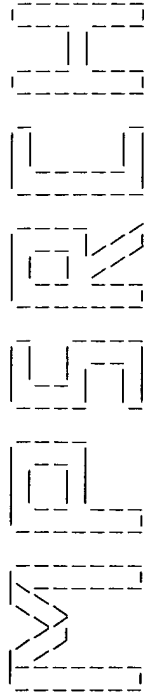
ORIGINAL SOURCE:

ORGANISM: Rhodococcus rhodochrous

CC STRAIN: J-1 (FERM BP-1478)  
SQ SEQUENCE 226 AA; 25201 MW; 263723 CN;  
Query Match 63.98; Score 39; DB 2; Length 226;  
Best Local Similarity 62.58; Pred. No. 2.91e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 17 KPESDEPV 24  
UY :|:|:|:|  
3 NPDSEPV 10

Search completed: Thu Oct 21 15:49:18 1999  
Job time : 8 secs.

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Oct 21 15:46:53 1999; MasPar time 3.17 Seconds  
Tabular output not generated. 126.356 Million cell updates/sec

Title: >US-09-040-485-9  
Description: (1-10) from US09040485.pep  
Perfect Score: 61  
Sequence: 1 QENPDSSEPV 10

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 20.418; Variance 24.141; scale 0.846

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	100.0	754	1	peptide-aspartate bet	1.14e-03
2	61	100.0	757	2	aspartyl beta-hydroxy	1.14e-03
3	48	78.7	1379	2	hepatocyte growth fac	1.55e+00
4	46	75.4	440	2	hypothetical protein	4.31e+00
5	43	70.5	184	2	early gland protein e	1.90e+01
6	43	70.5	278	1	tropoin T, slow skel	1.90e+01
7	43	70.5	395	2	hypothetical 44.4k pr	1.90e+01
8	43	70.5	1218	2	probable transport pr	1.90e+01
9	43	70.5	1390	1	hepatocyte growth fac	1.90e+01
10	43	70.5	1816	1	laminin alpha-4 chain	1.90e+01
11	42	68.9	167	2	hypothetical protein	3.08e+01
12	42	68.9	607	2	80K protein (allele C	3.08e+01
13	42	68.9	622	1	cell wall assembly re	3.08e+01
14	42	68.9	634	2	calcium binding PW29	3.08e+01
15	42	68.9	1214	2	probable finger prote	3.08e+01
16	41	67.2	200	2	conserved hypothetical	4.93e+01
17	41	67.2	244	2	SCS2 protein - yeast	4.93e+01
18	41	67.2	311	2	synecan-1 precursor	4.93e+01
19	41	67.2	313	2	synecan core protein	4.93e+01
20	41	67.2	338	2	hypothetical protein	4.93e+01
21	41	67.2	411	2	DNA binding protein b	4.93e+01
22	41	67.2	539	2	probable ctp syntheta	4.93e+01
23	41	67.2	539	2	CTP synthetase homolo	4.93e+01

24 41 67.2 727 2 A56879 diacylglycerol kinase 4.93e+01  
25 41 67.2 938 1 Q0B524 nuclear antigen ENA- 4.93e+01  
26 41 67.2 946 2 S27921 nuclear antigen ENA- 4.93e+01  
27 41 67.2 1857 2 S01787 fatty-acid synthase ( 7.83e+01  
28 40 65.6 151 2 I38243 zona binding protein 7.83e+01  
29 40 65.6 215 2 JC1133 alpha-s1-casein precu 7.83e+01  
30 40 65.6 319 2 S46422 F37A4.5 protein - Cae 7.83e+01  
31 40 65.6 355 2 A44851 3-isopropylmalate deh 7.83e+01  
32 40 65.6 333 2 S60465 dom-3 protein - Caeno 7.83e+01  
33 40 65.6 413 2 S39195 levanucrase - Erwini 7.83e+01  
34 40 65.6 439 2 S72829 trehalose-6-phosphati 7.83e+01  
35 40 65.6 459 2 B44498 radial spoke protein 7.83e+01  
36 40 65.6 472 1 A47402 fatty acid binding/tr 7.83e+01  
37 40 65.6 472 2 I49590 CD36 antigen - mouse 7.83e+01  
38 40 65.6 484 1 KIEEXY xylulokinase (EC 2.7. 7.83e+01  
39 40 65.6 685 2 S64158 hypothetical protein 7.83e+01  
40 40 65.6 735 2 A45033 myelin transcription 7.83e+01  
41 40 65.6 759 2 G69258 ATP-dependent RNA hel 7.83e+01  
42 40 65.6 774 2 JC6095 hepatocyte nuclear fa 7.83e+01  
43 40 65.6 1056 2 A53767 tracheobronchial muc 7.83e+01  
44 40 65.6 1088 2 B34106 protein kinase (EC 2. 7.83e+01  
45 40 65.6 1350 2 G36793 hypothetical protein 7.83e+01

## ALIGNMENTS

RESULT 1  
ENTRY BABOH #type complete  
TITLE peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine  
ALTERNATE\_NAMES aspartyl (asparaginyl) beta-hydroxylase  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 31-Dec-1993 #sequence\_revision 10-Feb-1995 #text\_change  
29-May-1998  
ACCESSIONS A42969; A39470; B39470; C39470; S27948  
REFERENCE A42969  
#authors Jia, S.; VanDusen, W.J.; Diehl, R.E.; Kohl, N.E.; Dixon, R.A.; Elliston, K.O.; Stern, A.M.; Friedman, P.A.  
#journal J. Biol. Chem. (1992) 267:14322-14327  
#title cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.  
#cross-references MUID:92332546  
#accession A42969  
#molecule\_type mRNA  
##residues 1-754 #label JIA  
##cross-references EMBL:M91213; NID:g162693; PID:g162694  
##experimental\_source brain  
#note sequence extracted from NCBI backbone (NCBIP:108534)  
REFERENCE A39470  
#authors Wang, Q.; VanDusen, W.J.; Petroski, C.J.; Garsky, V.M.; Stern, A.M.; Friedman, P.A.  
#journal J. Biol. Chem. (1991) 266:14004-14010  
#title Bovine liver aspartyl beta-hydroxylase. Purification and characterization.  
#cross-references MUID:91310689  
#accession A39470  
#molecule\_type protein  
##residues 289-328 #label WAN  
#accession B39470  
#molecule\_type protein  
##residues 615-'X', 617-630, 'XX', 633-634, 'X', 636, 'XX', 639-641  
#accession C39470  
#molecule\_type protein  
##residues 311-347, 'X', 349, 'X', 351-373, 'X', 375-379, 'X', 381-382  
#label WA3

COMMENT This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the peptidyl-aspartate substrate converts alpha-ketoglutarate to succinate and releases carbon dioxide.  
COMMENT Aspartic acid and asparagine residues in the EGF homology domain of certain plasma proteins serve as the peptidyl-aspartate substrate.  
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology

```

KEYWORDS      glycoprotein; oxidoreductase; transmembrane protein
FEATURE
2-56          #domain intracellular #status predicted #label INC\
57-78          #domain transmembrane #status predicted #label TRM\
289-754        #product peptide-aspartate beta-dioxygenase, 56K form
311-754        #product peptide-aspartate beta-dioxygenase, 52K form
337-370        #domain tetratricopeptide repeat homology #label TT1\
371-404        #domain tetratricopeptide repeat homology #label TT2\
13,96,466,702  #binding_site carbohydrate (Asn) (covalent) #status
                predicted
SUMMARY      #length 754 #molecular-weight 84998 #checksum 9667

Query Match      100.0%; Score 61; DB 1; Length 754;
Best Local Similarity 100.0%; Pred. No. 1.14e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 248 QENPDSSEPV 257
QY 1 QENPDSSEPV 10

RESULT 2
ENTRY      I38423      #type complete
TITLE      aspartyl beta-hydroxylase - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       29-May-1998 #sequence_revision 29-May-1998 #text_change
10-Jul-1998
ACCESSIONS I38423
REFERENCE   I38423
#authors   Koriath, F.; Gieffers, C.; Frey, J.
#journal   Gene (1994) 150:395-399
#title     Cloning and characterization of the human gene encoding
            aspartyl beta-hydroxylase.
#cross-references MUID:95121937
#accession  I38423
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-757 #label RES
#cross-references EMBL:U03109; NID:g458032
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase;
                tetratricopeptide repeat homology
FEATURE
54-75        #domain transmembrane #status predicted #label TRM
SUMMARY      #length 757 #molecular-weight 85498 #checksum 2143

Query Match      100.0%; Score 61; DB 2; Length 757;
Best Local Similarity 100.0%; Pred. No. 1.14e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 233 QENPDSSEPV 242
QY 1 QENPDSSEPV 10

RESULT 3
ENTRY      S01254      #type complete
TITLE      hepatocyte growth factor receptor precursor - mouse
CONTAINS   protein-tyrosine kinase (EC 2.7.1.112) met
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
08-Sep-1997
ACCESSIONS S01254; JH0115; A45453
REFERENCE   S01254
#authors   Chan, A.M.L.; King, H.W.S.; Deakin, E.A.; Tempest, P.R.;
            Hilkens, J.; Kroezen, V.; Edwards, D.R.; Willis, A.J.;
            Brookes, P.; Cooper, C.S.
#journal   Oncogene (1988) 2:593-599
#title     Characterization of the mouse met proto-oncogene.
#cross-references MUID:88262253
#accession  S01254
#molecule_type mRNA

```

```

#residues      1-1379 #label CHA
#cross-references EMBL:Y00671; NID:g53058; PID:g53059
REFERENCE      JH0112
#authors       Wilks, A.F.; Kurban, R.R.; Hovens, C.M.; Ralph, S.J.
#journal        Gene (1989) 85:67-74
#title         The application of the polymerase chain reaction to cloning
                members of the protein tyrosine kinase family.
#cross-references MUID:90152381
#accession     JH0115
#molecule_type mRNA
#residues      '1',1200-1254,'R',1256-1360,'T',1262-1368 #label WIL
#experimental_source hemopoietic cell
#note          the authors translated the codon ACG for residue 1261 as
                Lys
REFERENCE      A45453
#authors       Weidner, K.M.; Sachs, M.; Birschmeier, W.
#journal        J. Cell Biol. (1993) 121:145-154
#title         The Met receptor tyrosine kinase transduces motility,
                proliferation, and morphogenic signals of scatter
                factor/hepatocyte growth factor in epithelial cells.
#cross-references MUID:93209981
#accession     A45453
#status         preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues      924-935 #label WEI
GENETICS      met
CLASSIFICATION #superfamily hepatocyte growth factor receptor; protein
                kinase homology
KEYWORDS      ATP; autophosphorylation; glycoprotein; phosphoprotein;
                phosphotransferase; proto-oncogene; receptor; transmembrane
                protein; tyrosine-specific protein kinase
FEATURE
1-24          #domain signal sequence #status predicted #label SIG\
25-929        #domain extracellular #status predicted #label EXT\
25-302        #product hepatocyte growth factor receptor alpha chain
                #status predicted #label ACH\
308-1379      #product hepatocyte growth factor receptor beta chain
                #status predicted #label BCH\
930-954       #domain transmembrane #status predicted #label TMW\
955-1379      #domain intracellular #status predicted #label INT\
1074-1342     #region protein kinase homology #label KIN\
1082-1090     #region protein kinase ATP-binding motif\
1108          #active_site Lys #status predicted\
1233          #binding_site phosphate (Tyr) (covalent) (by
                autophosphorylation) #status predicted
SUMMARY      #length 1379 #molecular-weight 153547 #checksum 6033

Query Match      78.7%; Score 48; DB 2; Length 1379;
Best Local Similarity 70.0%; Pred. No. 1.55e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 347 QSKPDSAEPV 356
QY 1 QENPDSSEPV 10

RESULT 4
ENTRY      B71293      #type complete
TITLE      hypothetical protein TP0693 - syphilis spirochete
ORGANISM   #formal_name Treponema pallidum subsp. pallidum #common_name
                syphilis spirochete
DATE       24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
17-Mar-1999
ACCESSIONS B71293
REFERENCE   A71250
#authors     Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.;
            Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton,
            R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod,
            M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson,
            D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,
            L.; Artlich, P.; Bowman, C.; Cotton, M.D.; Fujii, C.;
            Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey,

```

```

#journal      L.; Weidman, J.; Smith, H.O.; Venter, J.C.
#title        Science (1998) 281:375-388
#description   Complete genome sequence of Treponema pallidum, the syphilis
#accession    B71293
#status       preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues     1-440 #label COL
#cross-references GB:AE001243; GB:AE000520; NID:g3322990; PID:g3323000
#experimental_source strain Nichols
GENETICS
#gene         TP0693
#summary      #length 440 #molecular-weight 47672 #checksum 4753
Query Match 75.4%; Score 46; DB 2; Length 440;
Best Local Similarity 70.0%; Pred. No. 4.31e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 244 QERPSSPEPV 253
Qy 1 QENPDSSEPV 10

RESULT 5
ENTRY   A61628 #type complete
TITLE   early gland protein egg-1 precursor - fruit fly (Drosophila
         virilis)
ORGANISM
#formal_name Drosophila virilis
DATE    08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change
ACCESSIONS
REFERENCE A61628; S57581
#authors   Thueroff, E.; Stoeven, S.; Kress, H.
#journal   Mech. Dev. (1992) 37:81-93
#title     Drosophila salivary glands exhibit a regional reprogramming
#accession A61628
#status    preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-184 #label THU
REFERENCE S57581
#authors   Stoeven, S.
#submission submitted to the EMBL Data Library, June 1995
#accession S57581
#status    preliminary
#molecule_type DNA
#residues 1-184 #label STO
#cross-references EMBL:249942; NID:g887425; PID:g887426
GENETICS
#gene      FlyBase:Dvir/Egpl
#cross-references FlyBase:FBgn0005594
#keywords   extracellular protein; salivary gland
#summary    #length 184 #molecular-weight 20567 #checksum 4761
Query Match 70.5%; Score 43; DB 2; Length 184;
Best Local Similarity 50.0%; Pred. No. 1.90e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 77 EESPEDEPV 86
Qy 1 QENPDSSEPV 10

RESULT 6
ENTRY   TP0693 #type complete
TITLE   troponin T, slow skeletal muscle - human
ALTERNATE_NAMES
ORGANISM troponin T1
#formal_name Homo sapiens #common_name man
DATE    20-Jun-1989 #sequence_revision 17-May-1996 #text_change
ACCESSIONS
REFERENCE A29783; B29783; JC2126; A57979

```

```

REFERENCE A29783
#authors   Gahlmann, R.; Troutt, A.B.; Wade, R.P.; Gunning, P.; Kedes,
#journal   J. Biol. Chem. (1987) 262:16122-16126
#title     Alternative splicing generates variants in important
#description functional domains of human slow skeletal troponin T.
#cross-references MUID:88058976
#accession A29783
#molecule_type mRNA
#residues 1-19, 'D', 21-278 #label GAH
#cross-references GB:M19309; NID:g339780; PID:g339781; GB:J03476
#experimental_source clone H22h
#accession B29783
#molecule_type mRNA
#residues 1-24, 36-204, 221-278 #label GA2
#cross-references GB:M19308; NID:g339782; PID:g339783; GB:J03476
#experimental_source clone M1
REFERENCE JC2126
#authors   Samson, F.; Mesnard, L.; Mihovilovic, M.; Potter, T.G.;
#journal   Mercadier, J.J.; Roses, A.D.; Gilbert, J.R.
#title     Biochem. Biophys. Res. Commun. (1994) 199:841-847
#description A new human slow skeletal troponin T (Tnts) mRNA isoform
#cross-references MUID:94183266
#accession JC2126
#molecule_type mRNA
#residues 1-204, 221-278 #label SAM
#cross-references GB:S69208; NID:g546020; PID:g546021
#accession A57979
#molecule_type mRNA
#residues 1-24, 36-204, 221-278 #label SA2
#cross-references GB:S69209; NID:g546022; PID:g546023
GENETICS
#gene      GDB:TNNI1
#map_position 19q13.4-19q13.4
#complex    troponin is a heterotrimer with one molecule each of troponin
           C (calcium binding component), troponin I (inhibitory
           component), and troponin T (tropomyosin-binding component)
FUNCTION
#description binds the troponin complex to tropomyosin; with tropomyosin
           mediates contraction of vertebrate striated muscle in
           response to calcium
#pathway     muscle contraction
CLASSIFICATION #superfamily troponin T
KEYWORDS       acetylated amino end; actin binding; alternative splicing;
           muscle contraction; phosphoprotein; skeletal muscle; thin
           filaments
FEATURE
2-278 #product troponin T, slow skeletal muscle splice form 1
#status predicted #label MAT1\
2-204,221-278 #product troponin T, slow skeletal muscle splice form 2
#status predicted #label MAT2\
2-24,36-204,221-278 #product troponin T, slow skeletal muscle splice form 3
#status predicted #label MAT3\
2 #modified site acetylated amino end (Ser) (in mature
  form) #status predicted\
2 #binding_site phosphate (Ser) (covalent) (by troponin T
  kinase) #status predicted\
177 #binding_site phosphate (Thr) (covalent) (by
    calmodulin-dependent kinase II) #status predicted
SUMMARY      #length 278 #molecular-weight 32948 #checksum 4551
Query Match 70.5%; Score 43; DB 1; Length 278;
Best Local Similarity 50.0%; Pred. No. 1.90e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 23 EAPPEPEPV 32
Qy 1 QENPDSSEPV 10

```

```

RESULT 7
ENTRY   JO0430      #type complete
TITLE   hypothetical 44.4K protein - Streptomyces fradiae transposon
        In4556
ORGANISM #formal_name Streptomyces fradiae
DATE     07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
        18-Jun-1993
ACCESSIONS JO0430
REFERENCE  JO0424
#authors  Stenieniak, D.R.; Slightom, J.L.; Chung, S.T.
#journal  Gene (1990) 86:1-9
#title    Nucleotide sequence of Streptomyces fradiae transposable
        element Tn4556: a class-II transposon related to Tn3.
#cross-references MUID:90185236
#accession JO0430
#molecule_type DNA
#residues  1-395 #label SIE
SUMMARY  #length 395 #molecular-weight 44379 #checksum 7125
        Query Match 70.5%; Score 43; DB 2; Length 395;
        Best Local Similarity 60.0%; Pred. No. 1.90e+01;
        Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 58 QERADGAEVP 67
QY 1 QENPDSSSEVP 10

RESULT 8
ENTRY   S38182      #type complete
TITLE   probable transport protein YKR103W - yeast (Saccharomyces
        cerevisiae)
ALTERNATE_NAMES Multidrug resistance protein homolog YKR103W
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     03-May-1994 #sequence_revision 03-May-1994 #text_change
        24-Jul-1998
ACCESSIONS S38182
REFERENCE  S38175
#authors  Gaillon, L.; Dujon, B.
#submission submitted to the Protein Sequence Database, March 1994
#accession S38182
#molecule_type DNA
#residues 1-1218 #label GAI
#cross-references EMBL:Z28328; NID:g486610; PID:g486611; MIPS:YKR103W
#experimental_source strain S288C
GENETICS  #map_position 11R
CLASSIFICATION #superfamily unassigned ATP-binding cassette proteins;
        ATP-binding cassette homology
        ATP; P-loop; transmembrane protein
KEYWORDS  #domain ATP-binding cassette homology #label ABC\
FEATURE   659-868 #region nucleotide-binding motif A (P-loop)\
        686-693 #binding_site ATP (Lys) #status predicted
        692
SUMMARY  #length 1218 #molecular-weight 137995 #checksum 1891
        Query Match 70.5%; Score 43; DB 2; Length 1218;
        Best Local Similarity 77.8%; Pred. No. 1.90e+01;
        Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 416 EENPDSSSEA 424
QY 1 QENPDSSSEP 9

RESULT 9
ENTRY   TVHUME      #type complete
TITLE   hepatocyte growth factor receptor precursor - human
CONTAINS protein-tyrosine kinase (BC 2.7.1.112) met
ORGANISM #formal_name Homo sapiens #common_name man
DATE     31-Mar-1991 #sequence_revision 30-Sep-1992 #text_change
        26-Feb-1999
ACCESSIONS A40175; A28303; A93749; A93369; A53761; I57632; A30008;

```

---

```

REFERENCE B24569
A40175
#authors  Giordano, S.
#submission submitted to the EMBL Data Library, November 1990
#accession A40175
#molecule_type mRNA
#residues 1-1390 #label GIO
#cross-references EMBL:X54559
REFERENCE  A28303
#authors  Park, M.; Dean, M.; Kaul, K.; Braun, M.J.; Gonda, M.A.; Vande
        Woude, G.
#journal  Proc. Natl. Acad. Sci. U.S.A. (1987) 84:6379-6383
#title    Sequence of MET protooncogene cDNA has features
        Characteristic of the tyrosine kinase family of
        growth-factor receptors.
#cross-references MUID:87317655
#accession A28303
#molecule_type mRNA
#residues 1-755, 'TWMKEPLNIYSLFCFAS', 756-1190, 'A', 1192-1390
        #label PAR
#cross-references GB:J02958; NID:g187558; PID:g307196
REFERENCE  A93749
#authors  Chan, A.M.L.; King, H.W.S.; Tempest, P.R.; Deakin, E.A.;
        Cooper, C.S.; Brookes, P.
#journal  Oncogene (1987) 1:229-233
#title    Primary structure of the met protein tyrosine kinase domain.
#cross-references MUID:88143699
#accession A93749
#molecule_type mRNA
#residues 'VNTRCQSLRLKLNKALTEKNKELEIAODRNIAIOQ',
        1273-1390 #label CHA
        'FTTKLELEAKRDLIRNRLSLELYT', 1010-1271, 'L',
        1273-1390 #label CHA
#cross-references GB:U08818; NID:g487741; PID:g487742
#note     this activated met oncogene is the product of gene
        rearrangement
REFERENCE  A93369
#authors  Dean, M.; Park, M.; Le Beau, M.M.; Robins, T.S.; Diaz, M.O.;
        Rowley, J.D.; Blair, D.G.; Vande Woude, G.F.
#journal  Nature (1985) 318:385-388
#title    The human met oncogene is related to the tyrosine kinase
        oncogenes.
#cross-references MUID:86065462
#accession A93369
#molecule_type DNA
#residues 1267-1390 #label DEA
#cross-references GB:M35074; NID:g187555; PID:g386868
REFERENCE  A53761
#authors  Gambartotta, G.; Pistoi, S.; Giordano, S.; Comoglio, P.M.;
        Santoro, C.
#journal  J. Biol. Chem. (1994) 269:12852-12857
#title    Structure and inducible regulation of the human MET promoter.
#cross-references MUID:94230365
#accession A53761
#molecule_type mRNA
#residues 1-14 #label GAM
REFERENCE  A40179
#authors  Ferracini, R.; Longati, P.; Naldini, L.; Vigna, E.; Comoglio,
        P.M.
#journal  J. Biol. Chem. (1991) 266:19558-19564
#title    Identification of the major autophosphorylation site of the
        Met/hepatocyte growth factor receptor tyrosine kinase.
#cross-references MUID:92011756
#contents annotation; autophosphorylation site
REFERENCE  I57632
#authors  Dean, M.; Park, M.; Vande Woude, G.F.
#journal  Mol. Cell. Biol. (1987) 7:921-924
#title    Characterization of the rearranged tpr-met oncogene
        breakpoint.
#cross-references MUID:87144265
#accession I57632
#status   translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 963-1009 #label RES

```



##cross-references GB:M15325; NID:g187531; PID:g187532  
 COMMENT The receptor is a dimer of disulfide-bonded 50K alpha and 145K beta chains that arise by cleavage of the precursor. Activity is regulated by phosphorylation of serine and tyrosine residues.

## GENETICS

#gene GDB:MET  
 ##cross-references GDB:120178; OMIM:164860  
 #map\_position 7q31-7q31  
 CLASSIFICATION #superfamily hepatocyte growth factor receptor; protein kinase homology  
 KEYWORDS ATP; autophosphorylation; glycoprotein; phosphoprotein; phosphotransferase; proto-oncogene; receptor; transmembrane protein; tyrosine-specific protein kinase

## FEATURE

1-24 #domain signal sequence #status predicted #label SIG\  
 25-303 #product hepatocyte growth factor receptor alpha chain #status predicted #label ALP\  
 308-1390 #product hepatocyte growth factor receptor beta chain #status predicted #label BET\  
 933-955 #domain transmembrane #status predicted #label TMN\  
 1076-1344 #domain protein kinase homology #label KIN\  
 1084-1092 #region protein kinase ATP-binding motif\  
 45,106,149,202,399, #binding\_site carbohydrate (Asn) (covalent) #status predicted\  
 405,635,785,930

1110 #active\_site Lys #status experimental\  
 1235 #binding\_site phosphate (Tyr) (covalent) (by autophosphorylation) #status experimental

## SUMMARY

#length 1390 #molecular-weight 155526 #checksum 2959

Query Match 70.5%; Score 43; DB 1; Length 1390;  
 Best Local Similarity 60.0%; Pred. No. 1.90e+01;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 348 QSKPDSAEPM 357

QY 1 QENPDSSEPV 10

## RESULT 10

ENTRY S68960 #type complete  
 TITLE laminin alpha-4 chain precursor - human  
 ALTERNATE\_NAMES laminin A4  
 ORGANISM Homo sapiens #common\_name man  
 DATE 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 26-Feb-1999  
 ACCESSIONS S68960; S65926; S49149; S40150; I53516  
 REFERENCE Richards, A.; Al-Imara, L.; Pope, F.M.  
 Eur. J. Biochem. (1996) 238:813-821  
 #journal The complete cDNA sequence of laminin alpha-4 and its relationship to the other human laminin alpha chains.  
 #title relationshp to the other human laminin alpha chains.  
 #cross-references MUID:96300249  
 #accession S68960  
 ##molecule\_type mRNA  
 ##residues 1-1816 ##label RIC  
 ##cross-references EMBL:X91171; NID:g1212962; PID:e198045; PID:g1212963  
 ##experimental\_source tissue type heart  
 REFERENCE I53516  
 #authors Iivanainen, A.; Sainio, K.; Sariola, H.; Tryggvason, K.  
 #journal FEBS Lett. (1995) 365:183-188  
 #title Primary structure and expression of a novel human laminin alpha-4 chain.  
 #cross-references MUID:95300971  
 #accession S65926  
 ##molecule\_type mRNA  
 ##residues 1-142, 'P', 144-177, 'F', 179-490, 'Y', 492-1056, 'P', 1058-1816  
 ##label IIV  
 ##cross-references EMBL:S78569; NID:g1042081; PID:g1042082  
 REFERENCE S49149  
 #authors Richards, A.J.; Al-Imara, L.; Carter, N.; Leversha, M.; Lloyd, J.C.; Pope, F.M.  
 #submission submitted to the EMBL Data Library, December 1993

#description Localisation of the gene (LAMA4) to chromosome 6q21 and isolation of a partial cDNA encoding a variant laminin A chain.  
 #accession S49149  
 ##molecule\_type mRNA  
 ##residues 236-1816 ##label RI2  
 ##cross-references EMBL:X76939; NID:g509805; PID:g509806  
 REFERENCE S40150  
 #authors Richards, A.J.; Al-Imara, L.; Carter, N.; Lloyd, J.C.; Pope, F.M.  
 #submission submitted to the EMBL Data Library, February 1993  
 #description Isolation of a partial cDNA encoding a protein homologous to laminin A. Assignment of the gene to chromosome 6.

#accession S40150  
 ##molecule\_type mRNA  
 ##residues 1403-1541, 'S', 1543-1816 ##label RI3  
 ##cross-references EMBL:X70904; NID:g437804; PID:g437805  
 GENETICS GDB:LAMA4; LAMA3  
 ##cross-references GDB:203904; OMIM:600133  
 #map\_position 6q21-6q21  
 COMPLEX Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin chain.

FUNCTION interact with cells and with other basement membrane proteins to promote differentiation, development, and cell migration  
 CLASSIFICATION #superfamily laminin alpha-4 chain; laminin G repeat homology; laminin-type EGF-like homology  
 KEYWORDS basement membrane; cell binding; coiled coil; extracellular matrix; glycoprotein; heptad repeat; heterotrimer

FEATURE 1-24 #domain signal sequence #status predicted #label SIG\  
 25-1816 #product laminin alpha-4 chain #status predicted #label MAT\  
 82-129 #domain laminin-type EGF-like homology #label LE1\  
 132-184 #domain laminin-type EGF-like homology #label LE2\  
 187-238 #domain laminin-type EGF-like homology #label LE3\  
 241-265 #domain laminin-type EGF-like homology #status atypical #label LE4\  
 717-719 #region cell attachment (R-G-D) motif

862-1003 #domain laminin G repeat homology #label LG1\  
 1068-1198 #domain laminin G repeat homology #label LG2\  
 1252-1367 #domain laminin G repeat homology #label LG3\  
 1488-1614 #domain laminin G repeat homology #label LG4\  
 1665-1789 #domain laminin G repeat homology #label LG5\  
 104,215,308,458, 524,550,571,574, 631,639,735,751, 754,780,803,1086, 1281,1359,1411

265,269 #binding\_site carbohydrate (Asn) (covalent) #status predicted\  
 #disulfide\_bonds interchain #status predicted  
 #length 1816 #molecular-weight 201882 #checksum 8148  
 SUMMARY

Query Match 70.5%; Score 43; DB 1; Length 1816;  
 Best Local Similarity 55.6%; Pred. No. 1.90e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 45 QDPENSEP 53

QY 1 QENPDSSEP 9

## RESULT 11

ENTRY E71016 #type complete  
 TITLE hypothetical protein PH1427 - Pyrococcus horikoshii  
 ORGANISM Pyrococcus horikoshii  
 DATE 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998  
 ACCESSIONS E71016  
 REFERENCE A71000  
 #authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;

```

Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:55-76
#journal
#title
Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MUID:98344137
#accession E71016
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-167, #label RAW
##cross-references GB:AP000006; NID:g3236133; PID:d1031476; PID:g3257850
##experimental_source strain OT3
##note
sequence replaced by GenBank
GENETICS
#gene PH1427
SUMMARY
#length 167 #molecular-weight 18831 #checksum 3158
Query Match 68.9%; Score 42; DB 2; Length 167;
Best Local Similarity 85.7%; Pred. No. 3.08e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 148 PESSEPV 154
I:|||||
Qy 4 PDSSEPV 10

RESULT 12
ENTRY S27776 #type complete
TITLE 80K protein (allele C1B) - Babesia bovis
ORGANISM #formal_name Babesia bovis
DATE 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
09-Sep-1997
ACCESSIONS S27776
REFERENCE S27776
#authors Dalrymple, B.P.; Peters, J.M.
#submission submitted to the EMBL Data Library, May 1992
#description Sequence of cDNA clones of a babesia bovis gene isolated
using sera from cattle vaccinated with a dextran sulphate
antigen fraction.
#accession S27776
##molecule_type mRNA
##residues 1-607 #label DAL
##cross-references EMBL:M93126; NID:g155862; PID:g155863
SUMMARY
#length 607 #molecular-weight 67129 #checksum 8353
Query Match 68.9%; Score 42; DB 2; Length 607;
Best Local Similarity 55.6%; Pred. No. 3.08e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 367 EEPAEPEPV 375
I:|||||
Qy 2 ENPDSEPV 10

RESULT 13
ENTRY A49344 #type complete
TITLE cell wall assembly regulatory protein SKN7 - yeast
(Saccharomyces cerevisiae)
ALTERNATE_NAMES oxidative stress response regulator POS9; protein YHR206w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
26-Feb-1999
ACCESSIONS A49344; S48987; S49986; S68114
REFERENCE A49344
#authors Brown, J.L.; North, S.; Bussey, H.
#journal J. Bacteriol. (1993) 175:6908-6915
#title SKN7, a yeast multicopy suppressor of a mutation affecting

```

```

cell wall beta-glucan assembly, encodes a product with
domains homologous to prokaryotic two-component regulators
and to heat shock transcription factors.
#cross-references MUID:94042854
#accession A49344
##molecule_type DNA
##residues 1-622 #label BRO
##cross-references GB:U00485; NID:g414418; PID:g414419
S46671
REFERENCE
#authors Macri, C.
#submission submitted to the EMBL Data Library, February 1994
#description The sequence of S. cerevisiae cosmid 9177.
#accession S48987
##molecule_type DNA
##residues 1-622 #label MAC
##cross-references EMBL:U00029; NID:g551322; PID:g458922; MIPS:YHR206w
S49986
REFERENCE
#authors Krems, B.; Charizanis, C.; Entian, K.D.
#submission submitted to the EMBL Data Library, November 1994
#description A protein (Pos9) similar to prokaryotic response regulators
is involved in oxidative stress in yeast.
#accession S49986
##molecule_type DNA
##residues 1-622 #label KRE
##cross-references EMBL:X83031; NID:g600027; PID:g600028
S68114
REFERENCE
#authors Krems, B.; Charizanis, C.; Entian, K.D.
#journal Curr. Genet. (1996) 29:327-334
#title The response regulator-like protein Pos9/Skn7 of
Saccharomyces cerevisiae is involved in oxidative stress
resistance
#cross-references MUID:96171515
#accession S68114
#status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-622 #label KRW
##cross-references EMBL:X83031; NID:g600027; PID:g600028
#note the nucleotide sequence was submitted to the EMBL Data
Library, November 1994
GENETICS
#gene SGD:SKN7; POS9
#cross-references SGD:S0001249; MIPS:YHR206w
#map_position 8R
CLASSIFICATION #superfamily cell wall assembly regulatory protein SKN7; HSF
DNA-binding domain homology; response regulator homology
DNA binding; leucine zipper; nucleus; phosphoprotein;
transcription regulation
KEYWORDS
FEATURE
87-194 #domain HSF DNA-binding domain homology #label HSF\
379-488 #domain response regulator homology #label RRR\
382-410 #region leucine zipper\
555-576 #region glutamine-rich\
427 #binding_site phosphate (Asp) (covalent) #status
predicted
SUMMARY
#length 622 #molecular-weight 69202 #checksum 256
Query Match 68.9%; Score 42; DB 1; Length 622;
Best Local Similarity 50.0%; Pred. No. 3.08e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 536 QDNPSSTTPV 545
I:|||||
Qy 1 QENPDSEPV 10

RESULT 14
ENTRY JC4248 #type complete
TITLE calcium binding PW29 protein - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 12-Oct-1995 #sequence_revision 08-Feb-1996 #text_change
07-Nov-1997
ACCESSIONS JC4248
REFERENCE JC4248

```

#authors Yu, S.; Ozawa, M.; Naved, A.F.; Miyauchi, T.; Muramatsu, H.;  
Muramatsu, T.  
#journal Cell Struct. Funct. (1995) 20:263-268  
#title cDNA cloning and sequence analysis of a novel calcium binding  
protein with oligoproline motif.  
#cross-references MUD:96098251  
#accession JC4248  
#molecule\_type mRNA  
#residues 1-634 #label YUS  
#cross-references DBJ:D49429; NID:9699609; PID:d1009004; PID:g1304155  
#experimental\_source F9 embryonal carcinoma cells

COMMENT This protein is a cytoplasmic calcium binding protein which lacks  
EF-hand motif, and is present in embryonal carcinoma cells. It  
plays important roles in regulation of cellular activities. This  
protein is rich in hydrophilic amino acids.  
calcium binding

KEYWORDS

FEATURE

528-547

SUMMARY #length 634 #molecular-weight 71893 #checksum 7243

Query Match 68.9%; Score 42; DB 2; Length 634;

Best Local Similarity 55.6%; Pred. No. 3.08e+01;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 276 DSPDSVDPV 284

::::: ::

QY 2 ENPDSSEPV 10

RESULT 15

ENTRY

S28499 #type complete

probable finger protein - rat

#formal\_name Rattus norvegicus #common\_name Norway rat

12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change

10-Sep-1997

ACCESSIONS S28499

REFERENCE S28499

#authors Hoeg, C.; Schalling, M.; Grunder-Brundell, E.; Daneholt, B.

#submission submitted to the EMBL Data Library, June 1991

#description Analysis of a murine germ cell-specific transcript that

encodes a putative zinc finger protein.

#accession S28499

#molecule\_type mRNA

#residues 1-1214 #label HOO

#cross-references EMBL:X59993; NID:g57503; PID:g57504

#experimental\_source strain Sprague Dawley

KEYWORDS DNA binding; zinc; zinc finger

SUMMARY #length 1214 #molecular-weight 135403 #checksum 4667

Query Match 68.9%; Score 42; DB 2; Length 1214;

Best Local Similarity 50.0%; Pred. No. 3.08e+01;

Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1084 QENPADHDPI 1093

::::: ::

QY 1 QENPDSSEPV 10

Search completed: Thu Oct 21 15:47:19 1999

Job time : 26 secs.

\*\*\*\*\*  
 M P S R C H (TM)  
 \*\*\*\*\*

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Thu Oct 21 15:47:36 1999; MasPar time 2.43 Seconds  
 Tabular output not generated. 116.421 Million cell updates/sec

Title: >US-09-040-485-9  
 Description: (1-10) from US09040485.pep  
 Perfect Score: 61  
 Sequence: 1 QENPDSSEPV 10

Scoring table: PAM 150  
 Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot37  
 1:swissprot

Statistics: Mean 21.064; Variance 21.885; scale 0.963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	61	100.0	754	1 ASPH_BOVIN	ASPARYL/ASPARAGINYL B 1.63e-04
2	61	100.0	757	1 ASPH_HUMAN	ASPARYL/ASPARAGINYL B 1.63e-04
3	48	78.7	1379	1 MET_MOUSE	HEPATOCTYTE GROWTH FACT 4.98e-01
4	45	73.8	304	1 CBP2_SIMV1	ZINC CARBOXYPEPTIDASE 2.69e+00
5	44	72.1	662	1 UL06_HSV6U	VIROIN PROTEIN U76 4.64e+00
6	43	70.5	277	1 TRH1_HUMAN	TROPONIN T, SLOW SKELE 7.93e+00
7	43	70.5	395	1 YH44_STRFR	HYPOTHETICAL 44.4 KD P 7.93e+00
8	43	70.5	475	1 TRB1_ECOLI	TRAB PROTEIN. 7.93e+00
9	43	70.5	995	1 AGAA_VIBS7	BETA-AGARASE A PRECURS 7.93e+00
10	43	70.5	1218	1 YK83_YEAST	PROBABLE ATP-DEPENDENT 7.93e+00
11	43	70.5	1390	1 MET_HUMAN	HEPATOCTYTE GROWTH FACT 7.93e+00
12	43	70.5	1816	1 LMA4_HUMAN	HEPATOCTYTE GROWTH FACT 7.93e+00
13	42	68.9	622	1 SKN7_YEAST	PUTATIVE TRANSCRIPTION 1.34e+01
14	42	68.9	1214	1 TSGA_RAT	TESTIS SPECIFIC PROTEIN 1.34e+01
15	41	67.2	244	1 SCS2_MOUSE	SCS2 PROTEIN. 2.25e+01
16	41	67.2	311	1 SDC1_MOUSE	SYNDECAN-1 PRECURSOR ( 2.25e+01
17	41	67.2	313	1 SDC1_RAT	SYNDECAN-1 PRECURSOR ( 2.25e+01
18	41	67.2	493	1 VG10_BMPD2	GENE 10 PROTEIN (GP10) 2.25e+01
19	41	67.2	539	1 PYRG_CHLTR	CTP SYNTHETASE (EC 6.3 2.25e+01
20	41	67.2	642	1 PHSA_STRAT	PHENOXAZINONE SYNTHASE 2.25e+01
21	41	67.2	662	1 UL06_HSV62	VIROIN PROTEIN U76. 2.25e+01
22	41	67.2	727	1 KDGA_RAT	DIACYLGLYCEROL KINASE. 2.25e+01
23	41	67.2	938	1 EBNA_EBV	EBNA-4 NUCLEAR PROTEIN 2.25e+01

24	41	67.2	1857	1 FAS2_PENPA	FATTY ACID SYNTHASE, S 2.25e+01
25	40	65.6	151	1 SP17_HUMAN	SPERM SURFACE PROTEIN 3.74e+01
26	40	65.6	163	1 SP17_PAPHA	SPERM SURFACE PROTEIN 3.74e+01
27	40	65.6	215	1 CAS1_RABIT	ALPHA CASEIN PRECURSOR 3.74e+01
28	40	65.6	319	1 YPT5_CAEEL	HYPOTHETICAL 35.8 KD P 3.74e+01
29	40	65.6	330	1 AAKG_MOUSE	5'-AMP-ACTIVATED PROTE 3.74e+01
30	40	65.6	331	1 AAKG_HUMAN	5'-AMP-ACTIVATED PROTE 3.74e+01
31	40	65.6	335	1 LEU3_SPLPL	3-ISOPROPYLMALATE DSHY 3.74e+01
32	40	65.6	339	1 PD12_SCHPO	PUTATIVE PROTEIN DISUL 3.74e+01
33	40	65.6	407	1 OZF_MOUSE	ZINC FINGER PROTEIN OZ 3.74e+01
34	40	65.6	415	1 SABC_ERWAM	LEVANSUCRASE (EC 2.4.1 3.74e+01
35	40	65.6	459	1 RSP6_CHLRE	FLAGELLAR RADIAL SPOKE 3.74e+01
36	40	65.6	471	1 CD36_RAT	PLATELET GLYCOPROTEIN 3.74e+01
37	40	65.6	471	1 CD36_MOUSE	PLATELET GLYCOPROTEIN 3.74e+01
38	40	65.6	484	1 XYLB_ECOLI	XYLOSE KINASE (EC 2. 3.74e+01
39	40	65.6	488	1 TY3H_ANGAN	TYROSINE 3-MONOOXYGENA 3.74e+01
40	40	65.6	586	1 PYRG_MYCTU	CTP SYNTHETASE (EC 6.3 3.74e+01
41	40	65.6	685	1 YGO4_YEAST	HYPOTHETICAL 78.1 KD P 3.74e+01
42	40	65.6	725	1 MYT1_HUMAN	MYELIN TRANSCRIPTION F 3.74e+01
43	40	65.6	785	1 YE15_CAEEL	HYPOTHETICAL 90.8 KD P 3.74e+01
44	40	65.6	1056	1 MUC5_HUMAN	TRACHEOBRONCHIAL MUCIN 3.74e+01
45	40	65.6	4568	1 DYHC_CAEEL	DYNEIN HEAVY CHAIN, CY 3.74e+01

## ALIGNMENTS

RESULT 1  
 ID ASPH\_BOVIN STANDARD; PRT; 754 AA.  
 AC Q28056;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ASPARYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE)  
 DE DE HYDROXYLASE)  
 DE DE OXYGENASE)  
 GN ASPH.  
 OS BOS TAURUS (BOVINE)  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=LIVER, AND BRAIN;  
 RX MEDLINE: 92332546.  
 RA JIA S., VANDUSEN W.J., DIEHL R.E., KOHL N.E., DIXON R.A.F.,  
 RA ELLISTON K.O., STERN A.M., FRIEDMAN P.A.;  
 RT "cdna cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase."  
 RT J. BIOL. CHEM. 267:14322-14327(1992).  
 RN [2]  
 RP SEQUENCE OF 289-385 AND 615-641.  
 RC TISSUE=LIVER;  
 RX MEDLINE: 91310689.  
 RA WANG Q., VANDUSEN W.J., PETROSKI C.J., GARSKY V.M., STERN A.M.,  
 RA FRIEDMAN P.A.;  
 RT "Bovine liver aspartyl beta-hydroxylase. Purification and characterization."  
 RT J. BIOL. CHEM. 266:14004-14010(1991).  
 CC -!- CERTAIN SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN PROTEINS.  
 CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -> PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).  
 CC -!- COFACTOR: IRON.  
 CC -!- SUBUNIT: MONOMER.  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.  
 CC -!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 289-754) OR 52 KD (AA 311-754) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
KW EMBL; M91213; G162694; -  
DR OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;  
KW ENDOPLASMIC RETICULUM.  
FT DOMAIN 1 57  
FT TRANSEM 58 78  
CYTOPLASMIC (POTENTIAL).  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT DOMAIN 79 754  
FT DOMAIN 9 12  
FT DOMAIN 14 21  
FT DOMAIN 318 328  
FT POLY-LYS.  
FT CARBOHYD 96 96  
FT POTENTIAL.  
FT CARBOHYD 466 466  
FT POTENTIAL.  
FT CARBOHYD 702 702  
SQ SEQUENCE 754 AA; 84998 MW; 608861B2 CRC32;

Query Match 100.0%; Score 61; DB 1; Length 754;  
Best Local Similarity 100.0%; Pred. No. 1.63e-04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 248 QENPDSSEPV 257  
| | | | | | | | | |  
Qy 1 QENPDSSEPV 10

RESULT 2  
ID ASPH\_HUMAN STANDARD; PRT; 757 AA.  
AC Q12797;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-  
DE HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-  
DE DIOXYGENASE).  
GN ASPH.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95121937.  
RA KORIOTH F., GIEFFERS C., FREY J.;  
RT "Cloning and characterization of the human gene encoding aspartyl  
RT beta-hydroxylase."  
RL GENE 150:395-399(1994).  
CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN  
CC CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF  
CC PROTEINS.  
CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) =  
CC PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).  
CC -!- COFACTOR: IRON.  
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC  
CC RETICULUM.  
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.  
CC -!- PFM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA  
CC 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY  
CC SIMILARITY).  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; U03109; G458032; -  
DR MIM; 600582; -  
KW OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;

KW ENDOPLASMIC RETICULUM.  
FT DOMAIN 1 54  
FT TRANSEM 55 75  
CYTOPLASMIC (POTENTIAL).  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT LUMENAL (POTENTIAL).  
FT POLY-SER.  
FT DOMAIN 13 20  
FT POLY-LYS.  
FT CARBOHYD 452 452  
FT POTENTIAL.  
FT CARBOHYD 705 705  
SQ SEQUENCE 757 AA; 85498 MW; A6AFC24 CRC32;  
Query Match 100.0%; Score 61; DB 1; Length 757;  
Best Local Similarity 100.0%; Pred. No. 1.63e-04;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 233 QENPDSSEPV 242  
| | | | | | | | | |  
Qy 1 QENPDSSEPV 10

RESULT 3  
ID MET\_MOUSE STANDARD; PRT; 1379 AA.  
AC P16056; Q62125;  
DT 01-APR-1990 (REL. 14, CREATED)  
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE  
DE TYROSINE KINASE) (EC 2.7.1.112) (HGF-SF RECEPTOR).  
GN MET.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88262253.  
RA CHAN A.M.L., KING H.W.S., DEAKIN E.A., TEMPEST P.R., HILKENS J.,  
RA KROEZEN V., EDWARDS D.R., WILLS A.J., BROOKES P., COOPER C.S.;  
RT "Characterization of the mouse met proto-oncogene."  
RL GENE 85:67-74(1989).  
RN ONCOGENE 2:593-599(1988).  
RN [2]  
RP SEQUENCE OF 1199-1270 FROM N.A.  
RX MEDLINE; 90152381.  
RA WILKS A.F., KURBAN R.R., HOVENS C.M., RALPH S.J.;  
RT "The application of the polymerase chain reaction to cloning members  
RT of the protein tyrosine kinase family."

CC -!- FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-  
CC PROTEIN KINASE ACTIVITY.  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
CC PROTEIN TYROSINE PHOSPHATE.  
CC -!- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA  
CC CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR  
CC (TRANSLOCATED PROMOTER) LOCUS OF CHROMOSOME 1 PRODUCES AN  
CC ONCOGENIC PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; Y00671; G53059; -  
DR EMBL; M33424; G200574; -  
DR PIR; S01254; S01254.  
DR MGD; MGI:96969; MET.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.

DR PFAM: PF00059; pkinase: 1.  
 DR HSP: Pli362; IFGI.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;  
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 1379 HEPATOCYTE GROWTH FACTOR RECEPTOR.  
 FT DOMAIN 25 931 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 932 954 POTENTIAL.  
 FT DOMAIN 955 1379 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1076 1343 PROTEIN KINASE.  
 FT SITE 306 307 CLEAVAGE (POTENTIAL).  
 FT NP\_BIND 1082 1090 ATP (BY SIMILARITY).  
 FT BINDING 1108 1108 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1202 1202 BY SIMILARITY.  
 FT MOD\_RES 1233 1233 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 45 45 POTENTIAL.  
 FT CARBOHYD 106 106 POTENTIAL.  
 FT CARBOHYD 201 201 POTENTIAL.  
 FT CARBOHYD 357 357 POTENTIAL.  
 FT CARBOHYD 398 398 POTENTIAL.  
 FT CARBOHYD 404 404 POTENTIAL.  
 FT CARBOHYD 606 606 POTENTIAL.  
 FT CARBOHYD 634 634 POTENTIAL.  
 FT CARBOHYD 784 784 POTENTIAL.  
 FT CARBOHYD 878 878 POTENTIAL.  
 FT CONFLICT 1199 1199 V -> I (IN REF. 2).  
 FT CONFLICT 1255 1255 T -> R (IN REF. 2).  
 FT CONFLICT 1261 1261 K -> T (IN REF. 2).  
 FT CONFLICT 1269 1270 VL -> IP (IN REF. 2).  
 FT SEQUENCE 1379 AA; 153548 MW; E1597F1A CRC32;  
 Query Match 78.7%; Score 48; DB 1; Length 1379;  
 Best Local Similarity 70.0%; Pred. No. 4.98e-01;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 347 QSKPDSSEPV 356  
 QY 1 QENPDSSEPV 10  
 RESULT 4  
 ID CBPZ-SINVI STANDARD; PRT: 304 AA.  
 AC P42788;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE ZINC CARBOXYPEPTIDASE (EC 3.4.17.-) (FRAGMENT).  
 OS SIMULIUM VITTATUM (BLACK FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
 OC PTERYGOTA; DIPTERA; NEMATOCERA; CHIRONOMOIDA; SIMULIIDAE; SIMULIUM.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-GUT.  
 RX MEDLINE; 94093864.  
 RA RAMOS A., MAHOWALD A., JACOBS-LORENA M.;  
 RT "gut-specific genes from the black fly Simulium vittatum encoding  
 RT trypsin-like and carboxypeptidase-like proteins.";  
 RL INSECT MOL. BIOL. 1:149-163(1993).  
 CC -!- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.  
 CC -!- TISSUE SPECIFICITY: GUT-SPECIFIC.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE  
 CC ZINC CARBOXYPEPTIDASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L08481; G161186;  
 DR PROSITE; PS00132; CARBOXYPEPT\_2N\_1; 1.  
 DR HSP: Pli362; IFGI.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;  
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 1379 HEPATOCYTE GROWTH FACTOR RECEPTOR.  
 FT DOMAIN 25 931 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 932 954 POTENTIAL.  
 FT DOMAIN 955 1379 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1076 1343 PROTEIN KINASE.  
 FT SITE 306 307 CLEAVAGE (POTENTIAL).  
 FT NP\_BIND 1082 1090 ATP (BY SIMILARITY).  
 FT BINDING 1108 1108 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1202 1202 BY SIMILARITY.  
 FT MOD\_RES 1233 1233 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 45 45 POTENTIAL.  
 FT CARBOHYD 106 106 POTENTIAL.  
 FT CARBOHYD 201 201 POTENTIAL.  
 FT CARBOHYD 357 357 POTENTIAL.  
 FT CARBOHYD 398 398 POTENTIAL.  
 FT CARBOHYD 404 404 POTENTIAL.  
 FT CARBOHYD 606 606 POTENTIAL.  
 FT CARBOHYD 634 634 POTENTIAL.  
 FT CARBOHYD 784 784 POTENTIAL.  
 FT CARBOHYD 878 878 POTENTIAL.  
 FT CONFLICT 1199 1199 V -> I (IN REF. 2).  
 FT CONFLICT 1255 1255 T -> R (IN REF. 2).  
 FT CONFLICT 1261 1261 K -> T (IN REF. 2).  
 FT CONFLICT 1269 1270 VL -> IP (IN REF. 2).  
 FT SEQUENCE 1379 AA; 153548 MW; E1597F1A CRC32;  
 Query Match 78.7%; Score 48; DB 1; Length 1379;  
 Best Local Similarity 70.0%; Pred. No. 4.98e-01;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 347 QSKPDSSEPV 356  
 QY 1 QENPDSSEPV 10  
 RESULT 4  
 ID CBPZ-SINVI STANDARD; PRT: 304 AA.  
 AC P42788;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE ZINC CARBOXYPEPTIDASE (EC 3.4.17.-) (FRAGMENT).  
 OS SIMULIUM VITTATUM (BLACK FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
 OC PTERYGOTA; DIPTERA; NEMATOCERA; CHIRONOMOIDA; SIMULIIDAE; SIMULIUM.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-GUT.  
 RX MEDLINE; 94093864.  
 RA RAMOS A., MAHOWALD A., JACOBS-LORENA M.;  
 RT "gut-specific genes from the black fly Simulium vittatum encoding  
 RT trypsin-like and carboxypeptidase-like proteins.";  
 RL INSECT MOL. BIOL. 1:149-163(1993).  
 CC -!- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.  
 CC -!- TISSUE SPECIFICITY: GUT-SPECIFIC.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE  
 CC ZINC CARBOXYPEPTIDASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L08481; G161186;  
 DR PROSITE; PS00132; CARBOXYPEPT\_2N\_1; 1.  
 DR HSP: Pli362; IFGI.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;  
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 1379 HEPATOCYTE GROWTH FACTOR RECEPTOR.  
 FT DOMAIN 25 931 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 932 954 POTENTIAL.  
 FT DOMAIN 955 1379 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1076 1343 PROTEIN KINASE.  
 FT SITE 306 307 CLEAVAGE (POTENTIAL).  
 FT NP\_BIND 1082 1090 ATP (BY SIMILARITY).  
 FT BINDING 1108 1108 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1202 1202 BY SIMILARITY.  
 FT MOD\_RES 1233 1233 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 45 45 POTENTIAL.  
 FT CARBOHYD 106 106 POTENTIAL.  
 FT CARBOHYD 201 201 POTENTIAL.  
 FT CARBOHYD 357 357 POTENTIAL.  
 FT CARBOHYD 398 398 POTENTIAL.  
 FT CARBOHYD 404 404 POTENTIAL.  
 FT CARBOHYD 606 606 POTENTIAL.  
 FT CARBOHYD 634 634 POTENTIAL.  
 FT CARBOHYD 784 784 POTENTIAL.  
 FT CARBOHYD 878 878 POTENTIAL.  
 FT CONFLICT 1199 1199 V -> I (IN REF. 2).  
 FT CONFLICT 1255 1255 T -> R (IN REF. 2).  
 FT CONFLICT 1261 1261 K -> T (IN REF. 2).  
 FT CONFLICT 1269 1270 VL -> IP (IN REF. 2).  
 FT SEQUENCE 1379 AA; 153548 MW; E1597F1A CRC32;  
 Query Match 78.7%; Score 48; DB 1; Length 1379;  
 Best Local Similarity 70.0%; Pred. No. 4.98e-01;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 347 QSKPDSSEPV 356  
 QY 1 QENPDSSEPV 10  
 RESULT 4  
 ID CBPZ-SINVI STANDARD; PRT: 304 AA.  
 AC P42788;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE ZINC CARBOXYPEPTIDASE (EC 3.4.17.-) (FRAGMENT).  
 OS SIMULIUM VITTATUM (

DR PFAM: PF00059; pkinase: 1.  
 DR HSP: Pli362; IFGI.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;  
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 1379 HEPATOCYTE GROWTH FACTOR RECEPTOR.  
 FT DOMAIN 25 931 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 932 954 POTENTIAL.  
 FT DOMAIN 955 1379 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1076 1343 PROTEIN KINASE.  
 FT SITE 306 307 CLEAVAGE (POTENTIAL).  
 FT NP\_BIND 1082 1090 ATP (BY SIMILARITY).  
 FT BINDING 1108 1108 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1202 1202 BY SIMILARITY.  
 FT MOD\_RES 1233 1233 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 45 45 POTENTIAL.  
 FT CARBOHYD 106 106 POTENTIAL.  
 FT CARBOHYD 201 201 POTENTIAL.  
 FT CARBOHYD 357 357 POTENTIAL.  
 FT CARBOHYD 398 398 POTENTIAL.  
 FT CARBOHYD 404 404 POTENTIAL.  
 FT CARBOHYD 606 606 POTENTIAL.  
 FT CARBOHYD 634 634 POTENTIAL.  
 FT CARBOHYD 784 784 POTENTIAL.  
 FT CARBOHYD 878 878 POTENTIAL.  
 FT CONFLICT 1199 1199 V -> I (IN REF. 2).  
 FT CONFLICT 1255 1255 T -> R (IN REF. 2).  
 FT CONFLICT 1261 1261 K -> T (IN REF. 2).  
 FT CONFLICT 1269 1270 VL -> IP (IN REF. 2).  
 FT SEQUENCE 1379 AA; 153548 MW; E1597F1A CRC32;  
 Query Match 78.7%; Score 48; DB 1; Length 1379;  
 Best Local Similarity 70.0%; Pred. No. 4.98e-01;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 347 QSKPDSSEPV 356  
 QY 1 QENPDSSEPV 10  
 RESULT 4  
 ID CBPZ-SINVI STANDARD; PRT: 304 AA.  
 AC P42788;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE ZINC CARBOXYPEPTIDASE (EC 3.4.17.-) (FRAGMENT).  
 OS SIMULIUM VITTATUM (BLACK FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
 OC PTERYGOTA; DIPTERA; NEMATOCERA; CHIRONOMOIDA; SIMULIIDAE; SIMULIUM.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-GUT.  
 RX MEDLINE; 94093864.  
 RA RAMOS A., MAHOWALD A., JACOBS-LORENA M.;  
 RT "gut-specific genes from the black fly Simulium vittatum encoding  
 RT trypsin-like and carboxypeptidase-like proteins.";  
 RL INSECT MOL. BIOL. 1:149-163(1993).  
 CC -!- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.  
 CC -!- TISSUE SPECIFICITY: GUT-SPECIFIC.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE  
 CC ZINC CARBOXYPEPTIDASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L08481; G161186;  
 DR PROSITE; PS00132; CARBOXYPEPT\_2N\_1; 1.  
 DR HSP: P00730; 4CPA.  
 KW HYDROLASE; CARBOXYPEPTIDASE; ZINC.  
 FT NON\_TER 1 1  
 FT METAL 58 58 ZINC (BY SIMILARITY).  
 FT METAL 61 61 ZINC (BY SIMILARITY).  
 FT METAL 184 184 ZINC (BY SIMILARITY).  
 FT ACT\_SITE 236 236 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).  
 FT DISULFID 135 148 BY SIMILARITY.  
 FT SEQUENCE 304 AA; 34849 MW; 9543CCAE CRC32;  
 Query Match 73.8%; Score 45; DB 1; Length 304;  
 Best Local Similarity 60.0%; Pred. No. 2.69e-00;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Db 17 QEHPEHVEPV 26  
 QY 1 QENPDSSEPV 10  
 RESULT 5  
 ID UL06\_HSV60 STANDARD; PRT: 662 AA.  
 AC P52453;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE VIRION PROTEIN U76.  
 GN U76 OR HDLF1.  
 OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).  
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;  
 OC BETAHERPESVIRINAE; ROSEOLVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95027704.  
 RA NICHOLAS J.;  
 RT "Nucleotide sequence analysis of a 21-kbp region of the genome of  
 RT human herpesvirus-6 containing homologues of human cytomegalovirus  
 RT major immediate-early and replication genes.";  
 RL VIROLOGY 204:758-750(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95266321.  
 RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,  
 RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;  
 RT "The DNA sequence of human herpesvirus-6: structure, coding content,  
 RT and genome evolution.";  
 RL VIROLOGY 209:29-51(1995).  
 CC -!- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA  
 CC PACKAGING.  
 CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6,  
 CC EHV-1 56, EBV BBRF1, HCMV UL104, AND VZV 54.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U13194; G662099;  
 DR EMBL; X83413; G834055;  
 SQ SEQUENCE 662 AA; 77234 MW; 1F2C2F67 CRC32;  
 Query Match 72.1%; Score 44; DB 1; Length 662;  
 Best Local Similarity 50.0%; Pred. No. 4.64e-00;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 612 QTHPENSEPT 621  
 QY 1 QENPDSSEPV 10

```

RESULT 6
ID TR1L_HUMAN STANDARD; PRT: 277 AA.
AC P13805;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TROPONIN T, SLOW SKELETAL MUSCLE ISOFORMS.
GN TNNT1 OR TNT.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA.
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE; 88058976.
RA GAHMANN R., TROUT A.B., WADE R.P., GUNNING P., REDES L.;
RT "Alternative splicing generates variants in important functional
RT domains of human slow skeletal troponin T.";
PL J. BIOL. CHEM. 262:16122-16126(1987).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE; 94183266.
RA SAMSON F., MESNARD L., MIHOVILOVIC M., POTTER T.G., MERCADIER J.-J.,
RA ROSES A.D., GILBERT J.R.;
RT "A new human slow skeletal troponin T (TnTs) mRNA isoform derived
RT from alternative splicing of a single gene.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 199:841-847(1994).
CC -!- FUNCTION: TROPONIN T IS THE TROPOMYOSIN-BINDING SUBUNIT OF
CC TROPONIN. THE THIN FILAMENT REGULATORY COMPLEX WHICH CONFERS
CC CALCIUM-SENSITIVITY TO STRIATED MUSCLE ACTOMYOSIN ATPASE ACTIVITY.
CC -!- ALTERNATIVE PRODUCTS: THE DIFFERENT ISOFORMS ARE GENERATED BY
CC ALTERNATIVE SPLICING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M19309; G339781; -
DR EMBL; M19308; G339783; -
DR EMBL; S69208; G546021; -
DR EMBL; S69209; G546023; -
DR PIR; A29783; A29783.
DR MIN; 191041; -
DR PFAM; PF00992; Troponin; 1.
KW MUSCLE PROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING;
KW MULTIGENE FAMILY.
FT INIT_MET 0
FT MOD_RES 1 1 PHOSPHORYLATION (BY CK2)
FT VARSPLIC 24 34 MISSING (IN SECOND ISOFORM).
FT VARSPLIC 204 219 MISSING (IN SECOND AND THIRD ISOFORMS).
FT CONFLICT 19 19 E -> D (IN REF. 1).
SQ SEQUENCE 277 AA; 32817 MW; B0685CC1 CRC32;

Query Match 70.5%; Score 43; DB 1; Length 277;
Best Local Similarity 50.0%; Pred. No. 7.93e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 22 EEAPEPEPV 31
Qy 1 QENPDSSEPV 10

RESULT 7
ID YT44_STRFR STANDARD; PRT: 395 AA.
AC P20188;
DT 01-FEB-1991 (REL. 17, CREATED)

```





[3] SEQUENCE OF 1010-1390 FROM N.A.  
RX MEDLINE; 88143699.  
RA CHAN A.M.L., KING H.W.S., TEMPEST P.R., DEAKIN E.A., COOPER C.S.,  
RA BROOKES P.;  
RT "Primary structure of the met protein tyrosine kinase domain."; ~  
RL ONCOGENE 1:229-233(1987).  
RN [4]  
RP SEQUENCE OF 1206-1264 FROM N.A.  
RX MEDLINE; 94067791.  
RA LEE S.T., STRUNK K.M., SPRITZ R.A.;  
RT "A survey of protein tyrosine kinase mRNAs expressed in normal human  
RT melanocytes.";  
RL ONCOGENE 8:3403-3410(1993).  
RN [5]  
RP SEQUENCE OF 1267-1390 FROM N.A.  
RX MEDLINE; 86085462.  
RA DEAN M., PARK M., LE BEAU M.M., ROBINS T.S., DIAZ M.O., ROWLEY J.D.,  
RA BLAIR D.G., VANDE WOUDE G.F.;  
RT "The human met oncogene is related to the tyrosine kinase oncogenes.";  
RL NATURE 318:385-388(1985).  
RN [6]  
RP SEQUENCE OF 1-754 FROM N.A.  
RX PAULEY A., ANDREWS S.;  
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBSJ DATA BANKS.  
RN [7]  
RP FUNCTION.  
RX MEDLINE; 91118019.  
RA BOTTARO D.P., RUBIN J.S., FALETTO D.L., CHAN A.M.-L., KMIECIK T.E.,  
RA VANDE WOUDE G.F., AARONSON S.A.;  
RT "Identification of the hepatocyte growth factor receptor as the c-met  
RT proto-oncogene product.";  
RL SCIENCE 251:802-804(1991).  
RN [8]  
RP PHOSPHORYLATION AT TYR-1235.  
RX MEDLINE; 92011756.  
RA FERRACINI R., LONGATI P., NALDINI L., VIGNA E., COMOGGIO P.M.;  
RT "Identification of the major autophosphorylation site of the  
RT Met/hepatocyte growth factor receptor tyrosine kinase.";  
RL J. BIOL. CHEM. 266:19558-19564(1991).  
RN [9]  
RP VARIANTS HPRC, AND VARIANT VAL-320.  
RX MEDLINE; 97285124.  
RA SCHMIDT L., DUH F.-M., CHEN F., KISHIDA T., GLENN G., CHOYKE P.,  
RA SCHERER S.W., ZHUANG Z., LUBENSKY I., DEAN M., ALLIKMETS R.,  
RA CHIDAMBARAM A., BERGERHEIM U.R., FELTIS J.T., CASADEVALL C.,  
RA ZAMARRON A., BERNUES M., RICHARD S., LIPS C.J.M., WALTHER M.M.,  
RA TSUI L.-C., GEIL L., ORCUTT M.L., STACKHOUSE T., LIPAN J., SLIFE L.,  
RA BRAUCH H., DECKER J., NIEHANS G., HUGHSON M.D., MOCH H., STORKEL S.,  
RA LERMAN M.I., LINEHAN W.M., ZBAR B.;  
RT "Germline and somatic mutations in the tyrosine kinase domain of the  
RT Met proto-oncogene in papillary renal carcinomas.";  
RL NAT. GENET. 16:68-73(1997).  
RN [10]  
RP FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-  
CC PROTEIN KINASE ACTIVITY.  
CC [11]  
CC CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
CC PROTEIN TYROSINE PHOSPHATE.  
CC [12]  
CC SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA  
CC CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.  
CC [13]  
CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC [14]  
CC DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR  
CC GENE PRODUCES AN ONCOGENIC PROTEIN.  
CC [15]  
CC DISEASE: DEFECTS IN MET ARE THE CAUSE OF HEREDITARY PAPILLARY  
CC RENAL CARCINOMA (HPRC). HPRC IS A FORM OF INHERITED KIDNEY CANCER  
CC CHARACTERIZED BY A PREDISPOSITION TO DEVELOP MULTIPLE, BILATERAL  
CC PAPILLARY RENAL TUMORS. THE PATTERN OF INHERITANCE IS CONSISTENT  
CC WITH AUTOSOMAL DOMINANT TRANSMISSION WITH REDUCED PENETRATION.  
CC [16]  
CC SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.  
CC [17]  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M35074; G386868; -;  
DR EMBL; X54559; -; NOT ANNOTATED\_CDS.  
DR EMBL; J02958; G307196; -;  
DR EMBL; AC002080; G2078456; -;  
DR PIR; A40175; TVHUME.  
DR MIM; 164860; -;  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PFAM; PF00069; pkinase; 1.  
DR HSP; F11362; IFGI.  
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;  
KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL;  
KW CHROMOSOMAL TRANSLOCATION; DISEASE MUTATION; POLYMORPHISM.  
FT SIGNAL 1 24  
FT CHAIN 25 1390 HEPATOCYTE GROWTH FACTOR RECEPTOR.  
FT DOMAIN 25 932 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 933 955 POTENTIAL.  
FT DOMAIN 956 1390 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 1078 1345 PROTEIN KINASE.  
FT NP\_BIND 1084 1092 ATP (BY SIMILARITY).  
FT BINDING 1110 1110 ATP (BY SIMILARITY).  
FT ACT\_SITE 1204 1204 BY SIMILARITY.  
FT SITE 307 308 CLEAVAGE (POTENTIAL).  
FT SITE 1009 1010 BREAKPOINT FOR TRANSLOCATION TO FORM  
FT TPR-MET ONCOGENE.  
FT MOD\_RES 1235 1235 PHOSPHORYLATION (AUTO-).  
FT CARBOHYD 45 45 POTENTIAL.  
FT CARBOHYD 106 106 POTENTIAL.  
FT CARBOHYD 149 149 POTENTIAL.  
FT CARBOHYD 202 202 POTENTIAL.  
FT CARBOHYD 399 399 POTENTIAL.  
FT CARBOHYD 405 405 POTENTIAL.  
FT CARBOHYD 607 607 POTENTIAL.  
FT CARBOHYD 635 635 POTENTIAL.  
FT CARBOHYD 785 785 POTENTIAL.  
FT CARBOHYD 879 879 POTENTIAL.  
FT CARBOHYD 930 930 POTENTIAL.  
FT VARIANT 320 320 A -> V.  
FT VARIANT 1131 1131 M -> T (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1188 1188 V -> L (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1195 1195 L -> V (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1220 1220 V -> I (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1228 1228 D -> N (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1228 1228 D -> H (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1230 1230 Y -> C (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1230 1230 Y -> H (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1250 1250 M -> T (IN HPRC; SOMATIC MUTATION).  
FT CONFLICT 755 755 S -> STWKKEPLNIVSLFCFAS (IN REF. 2).  
FT CONFLICT 1191 1191 G -> A (IN REF. 2).  
SQ SEQUENCE 1390 AA; 155526 MW; 650992C2 CRC32;  
  
Query Match 70.5%; Score 43; DB 1; Length 1390;  
Best Local Similarity 60.0%; Pred. No. 7.93e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Db 348 QSKPDSAEPM 357  
QY 1 QENPDSSEPV 10  
| :|||:|:  
| :|||:|:  
  
RESULT 12  
ID LMA4\_HUMAN STANDARD; PRT; 1816 AA.  
AC Q16363; Q15335; Q14735;  
DT 01-NOV-1997 (REL. 35; CREATED)  
DT 01-NOV-1997 (REL. 35; LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36; LAST ANNOTATION UPDATE)  
DE LAMININ ALPHA-4 CHAIN PRECURSOR.  
GN LAMA4.



OC RN  
RP  
PX MEDLINE; 94042854.  
TX "SKN7, a yeast multicopy suppressor of a mutation affecting cell wall  
RT beta-glucan assembly, encodes a product with domains homologous to  
RT prokaryotic two-component regulators and to heat shock transcription  
RT factors.";  
RL J. BACTERIOL. 175:6908-6915(1993).  
RN [2]  
RP  
RP SEQUENCE FROM N.A.  
RA KREMS B., CHARIZANIS C., ENTIAN K.-D.;  
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / AB972;  
RX MEDLINE: 94378003.  
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,  
RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,  
RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,  
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,  
RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,  
RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,  
RA VAUDIN M.;  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
VIII.";  
RL SCIENCE 265:2077-2082(1994).  
RN [4]  
RP  
RP FUNCTION, AND MUTAGENESIS.  
RX MEDLINE: 95045411.  
RA BROWN J.L., BUSSEY H., STEWART R.C.;  
RT "Yeast Skn7p functions in a eukaryotic two-component regulatory  
RT pathway.";  
RL EMBO J. 13:5186-5194(1994).  
CC  
CC -!- FUNCTION: INVOLVED IN OXIDATIVE STRESS. TRANSCRIPTION FACTOR THAT  
CC ACTS IN PARALLEL WITH THE PKC1 CASCADE TO REGULATE GROWTH AT THE  
CC CELL SURFACE.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: BELONGS TO THE HSF FAMILY.  
CC -!- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U00485; G414419; -  
CC EMBL; X83031; G600028; -  
CC EMBL; U00029; G458922; -  
CC PIR; A49344; A49344.  
CC PIR; S48987; S48987.  
CC SGD; L0001908; SKN7.  
CC PROSITE; PS00434; HSF\_DOMAIN; 1.  
CC PFAM; PF00072; response\_reg; 1.  
CC PFAM; PF00447; HSF\_DNA-Bind; 1.  
CC HSP; P22121; ZHTS.  
CC TRASP; P22121; ZHTS.  
KW TRANSCRIPTION REGULATION; SENSORY TRANSDUCTION; NUCLEAR PROTEIN;  
KW DNA-BINDING; PHOSPHORYLATION.  
KW DNA\_BIND 86 190 BY SIMILARITY.  
FT MOD\_RES 427 427 PHOSPHORYLATION (PROBABLE).  
FT MUTAGEN 427 427 D->N: DIMINISHED ACTIVITY.  
FT MUTAGEN 427 427 D->E: AUGMENTED ACTIVITY.  
SQ SEQUENCE 622 AA; 69202 MW; 4E506931 CRC32;  
  
Query Match 68.9%; Score 42; DB 1; Length 622;  
Best Local Similarity 50.0%; Pred. No. 1.34e+01;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
Db 1084 QENPADHPI 1093  
QY 1 QENPDSSEPV 10  
|||||:|:  
1 QENPDSSEPV 10  
  
Query Match 68.9%; Score 42; DB 1; Length 1214;  
Best Local Similarity 50.0%; Pred. No. 1.34e+01;  
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Db 1084 QENPADHPI 1093  
QY 1 QENPDSSEPV 10  
|||||:|:  
1 QENPDSSEPV 10  
  
RESULT 15  
ID SCS2\_YEAST STANDARD; PRT; 244 AA.  
AC P40075;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE SCS2 PROTEIN.  
GN SCS2 OR YER120W.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA NIKAWA J.-I., MURAKAMI A., ESUMI E., HOSAKA K.;  
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / AB972;  
RA DIERTRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,  
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,  
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,

RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,  
 RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,  
 RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,  
 RA TAYLOR P., WEI Y., YELTON M., BOTSSTEIN D., DAVIS R.W.;  
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -!- FUNCTION: CAN SUPPRESS AN INOSITOL AUXOTROPHIC MUTANT AND A  
 CC CHOLINE SENSITIVE MUTANT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D44493; G624933; -;  
 DR EMBL; U18916; G603359; -;  
 DR SGD; L0002629; SCS2.  
 DR PFAM; PF00635; MSP\_domain; 1.  
 SQ SEQUENCE 244 AA; 26925 MW; DA1802FD CRC32;

Query Match 67.2%; Score 41; DB 1; Length 244;  
 Best Local Similarity 60.0%; Pred. No. 2.25e+01;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 140 QENKVEPV 149  
 QY 1 QENPDSSEPV 10

Search completed: Thu Oct 21 15:47:46 1999  
 Job time : 10 secs.

\*\*\*\*\*

W P E R E F (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 21 15:48:03 1999; Maspar time 4.47 Seconds  
Tabular output not generated. 122.196 Million cell updates/sec

Title: >US-09-040-485-9  
Description: (1-10) from US09040485.pep  
Perfect Score: 61  
Sequence: 1 QENPDSEPV 10

Scoring table: PAM 150  
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sprembl9  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organellae  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 20.182; Variance 22.725; scale 0.888

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	49	80.3	374	13	O42322	D4B DOPAMINE RECEPTOR. 6.83e-01
2	46	75.4	344	5	O62447	Y43F4B.1 PROTEIN. 3.50e+00
3	46	75.4	440	2	O83691	HYPOTHETICAL 47.7 KD P 3.50e+00
4	44	72.1	590	6	O02665	SODIUM-D-GLUCOSE COIRA 1.00e+01
5	44	72.1	628	2	O67990	METALLOPROTEASE (FRAGM 1.00e+01
6	44	72.1	1262	5	O20771	F54D5.5. 1.00e+01
7	43	70.5	105	4	O14421	GLYCOPHORIN MZ (FRAGME 1.88e+01
8	43	70.5	129	4	O99737	LAMININ ALPHA 4 CHAIN. 1.88e+01
9	43	70.5	158	5	O20474	F46F6.3 PROTEIN. 1.88e+01
10	43	70.5	184	5	O24743	EGP-1 PRECURSOR. 1.88e+01
11	43	70.5	262	11	O88346	SLOW SKELETAL MUSCLE T 1.88e+01
12	43	70.5	480	5	O27033	MEMBRANE PROTEIN. 1.88e+01
13	43	70.5	1217	11	O88573	PROTO-ONCOGENE A24. 1.88e+01
14	43	70.5	1382	11	P97579	HEPATOCYTE GROWTH FACT 1.88e+01
15	43	70.5	1382	11	P97523	HGF RECEPTOR PRECURSOR 1.88e+01
16	42	68.9	81	4	O99483	GLYCOSYLTRANSFERASE (F 2.78e+01
17	42	68.9	139	5	P91281	COSMID F27C1. 2.78e+01
18	42	68.9	156	10	O39754	CRP1. 2.78e+01
19	42	68.9	167	1	O59097	167AA LONG HYPOTHETICA 2.78e+01
20	42	68.9	345	14	O56987	COAT PROTEIN. 2.78e+01

21	42	68.9	607	5	Q17112	80 KDA PROTEIN. 2.78e+01
22	42	68.9	629	13	O93310	14S COHESIN RAD21 SUBU 2.78e+01
23	42	68.9	631	4	O60216	DOUBLE-STRAND-BREAK RE 2.78e+01
24	42	68.9	634	11	Q61550	DOUBLE-STRAND-BREAK RE 2.78e+01
25	42	68.9	1211	11	O35233	HOMOLOG OF HUMAN MLT2 2.78e+01
26	41	67.2	140	2	O46022	XYLX GENE (FRAGMENT). 4.58e+01
27	41	67.2	157	6	O62841	CATHELICIDIN PRECURSOR 4.58e+01
28	41	67.2	200	2	O67307	HYPOTHETICAL 23.1 KD P 4.58e+01
29	41	67.2	205	5	Q17564	COIF6.8 PROTEIN. 4.58e+01
30	41	67.2	211	2	O07446	INSERTION ELEMENT IS14 4.58e+01
31	41	67.2	283	5	Q26940	HYPOTHETICAL P284 PROT 4.58e+01
32	41	67.2	338	1	O28769	HYPOTHETICAL 37.4 KD P 4.58e+01
33	41	67.2	411	10	O22763	PUTATIVE BZIP-LIKE DNA 4.58e+01
34	41	67.2	482	2	O52762	CATALASE (EC 1.11.1.6) 4.58e+01
35	41	67.2	484	2	P77939	CATALASE (EC 1.11.1.6) 4.58e+01
36	41	67.2	539	2	O84186	CTP SYNTHETASE. 4.58e+01
37	41	67.2	625	11	O35305	RECEPTOR ACTIVATOR OF 4.58e+01
38	41	67.2	662	14	O57138	SIMILAR TO HV6A U76. 4.58e+01
39	41	67.2	730	11	O88673	ALPHA DIACYLGLYCEROL K 4.58e+01
40	41	67.2	919	10	O24375	ALPHA-GLUCOSIDASE (EC 4.58e+01
41	41	67.2	946	14	O69139	NUCLEAR ANTIGEN EBNA-3 4.58e+01
42	41	67.2	991	5	O18152	T28D6.4 PROTEIN. 4.58e+01
43	41	67.2	1186	14	Q90061	NUCLEAR ANTIGEN-3B {EX 4.58e+01
44	41	67.2	1655	5	O44498	F15E6.1 PROTEIN. 4.58e+01
45	40	65.6	1078	11	O08995	MYELIN TRANSCRIPTION F 7.46e+01

## ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	374 AA.
ID O42322			
AC O42322			
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)			
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE D4B DOPAMINE RECEPTOR.			
OS CYPRINUS CARPIO (COMMON CARP).			
OC EUKARYOTA; METAQOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;			
OC TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;			
OC CYPRINIDAE; CYPRININAE; CYPRINUS.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE-RETINA;			
RA HIRANO J., ARCHER S.N., DJANGOZ M.B.A.;			
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.			
DR EMBL; Y14633; E334823; -			
DR PFAM; PF00001; 7tm_1; 1.			
SQ SEQUENCE 374 AA; 42004 MW; 185EF905 CRC32;			

Query Match 80.3%; Score 49; DB 13; Length 374;  
Best Local Similarity 60.0%; Pred. No. 6.83e-01;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 255 EQDPDSEPV 264	PRELIMINARY;	PRT;	344 AA.
Qy 1 QENPDSEPV 10			
	:::		

RESULT 2	PRELIMINARY;	PRT;	344 AA.
ID O62447			
AC O62447			
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)			
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)			
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)			
DE Y43F4B.1 PROTEIN.			
GN Y43F4B.1.			
OS CAENORHABDITIS ELEGANS.			
OC EUKARYOTA; METAQOA; NEMATODA; SECNENTEA; RHABDITIA; RHABDITIDA;			
OC RHABDITINA; RHABDITOIDEA; RHABDITINAE; PELODERINAE; CAENORHABDITIS.			
RN [1]			
RP SEQUENCE FROM N.A.			
RA MATTHEWS L.;			
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.			

```

RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA CRAXTON M., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RI NATURE 368:32-38(1994).
DR EMBL; AL021481; E1350408; -.
SQ SEQUENCE 344 AA; 38855 MW; FFAE32D6 CRC32;

Query Match 75.4%; Score 46; DB 5; Length 344;
Best Local Similarity 60.0%; Pred. No. 3.50e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 260 QENPEAPERY 269
QY 1 QENPDSEPV 10
|||||:|
|:|:|:|:|

RESULT 3
ID O83691 PRELIMINARY; PRT; 440 AA.
AC O83691:
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL 47.7 KD PROTEIN.
GN TP0693
OS TREPONEMA PALLIDUM.
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 9832770.
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RT "Complete Genome Sequence of Treponema pallidum, the Syphilis
RT Spirochete.";
RL SCIENCE 281:375-388(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE001243; G3323000; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 440 AA; 47672 MW; 34D6A5B3 CRC32;

Query Match 75.4%; Score 46; DB 2; Length 440;
Best Local Similarity 70.0%; Pred. No. 3.50e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 244 QERPPSPERY 253
QY 1 QENPDSEPV 10
|||||:|
|:|:|:|:|

```

```

RESULT 4
ID O02665 PRELIMINARY; PRT; 590 AA.
AC O02665:
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE SODIUM-D-GLUCOSE COTRANSPORTER.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RA REINEHARDT J., GAMBARIAN S., VEYLL M., KOEPESELL H.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X82876; E81419; -.
SQ SEQUENCE 590 AA; 62110 MW; 4686760C CRC32;

Query Match 72.1%; Score 44; DB 6; Length 590;
Best Local Similarity 75.0%; Pred. No. 1.00e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 43 DNPDPSTEP 50
QY 2 ENPDSEPV 9
|||||:|
|:|:|:|:|

RESULT 5
ID O67990 PRELIMINARY; PRT; 628 AA.
AC O67990:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE METALLOPROTEASE (FRAGMENT).
GN VMC.
OS VIBRIO MIMICUS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIONACEAE; VIBRIO.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC33653;
RA KONG I.S.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF004832; G3142333; -.
KW PROTEASE; METALLOPROTEASE.
FT NON_TER 1
SQ SEQUENCE 628 AA; 71233 MW; 58D32E50 CRC32;

Query Match 72.1%; Score 44; DB 2; Length 628;
Best Local Similarity 87.5%; Pred. No. 1.00e+01;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 568 ENPDSEPV 575
QY 2 ENPDSEPV 9
|||||:|
|:|:|:|:|

RESULT 6
ID Q20771 PRELIMINARY; PRT; 1262 AA.
AC Q20771:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F54D5.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA COLES L., MATTHEWS L.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.

```

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL NATURE 368:32-38(1994).  
DR EMBL: 266513; G104133;  
SQ SEQUENCE 1262 AA; 144630 MW; 7C2D2904 CRC32;

Query Match 72.1%; Score 44; DB 5; Length 1262;

Best Local Similarity 30.0%; Pred. No. 1.00e+01;

Matches 3; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Db 1120 QDSPQTADPI 1129

||||:|

QY 1 QENPDSSEP 10

RESULT 7 PRELIMINARY; PRT; 105 AA.

ID Q14421

AC Q14421

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE GLYCOPHORIN MZ (FRAGMENT).

GN GIPA.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

OC CATARRHINI; HOMINIDAE; HOMO.

OC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 93186803.

RA HUANG C.H., REID M.E., BLUMENFELD O.O.;

RT "Exon skipping caused by DNA recombination that introduces a

RT defective donor splice site into the human glycoporin A gene.";

RL J. BIOL. CHEM. 268:4945-4952(1993).

DR EMBL: L07253; G183325; .

DR PROSITE; PS00312; GLYCOPHORIN\_A; 1.

DR PFAM; PF01102; Glycophorin\_A; 1.

KW ALTERNATIVE SPLICING.

FT CHAIN 20 >105 POTENTIAL.

FT NON-TER 105 105

SQ SEQUENCE 105 AA; 11086 MW; 2FAB4CCC CRC32;

Query Match 70.5%; Score 43; DB 4; Length 105;

Best Local Similarity 62.5%; Pred. No. 1.68e+01;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 98 ENPETS DP 105

||||:|

QY 2 ENPDSSEP 9

RESULT 8

ID Q99737

AC Q99737

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE LAMININ ALPHA 4 CHAIN.

GN LAM4\*-1.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

OC CATARRHINI; HOMINIDAE; HOMO.

OC [1]

RP SEQUENCE FROM N.A.

RA XIAO S., LUX M.L., REEVES R., HUDSON T.J., FLETCHER J.A.;  
RL AM. J. PATHOL. 0:0-0(0).  
DR EMBL: U77706; G1684837;  
SQ SEQUENCE 129 AA; 13462 MW; 05562347 CRC32;

Query Match 70.5%; Score 43; DB 4; Length 129;

Best Local Similarity 55.6%; Pred. No. 1.68e+01;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 45 QDPPESEP 53

||||:|

QY 1 QENPDSSEP 9

RESULT 9

ID Q20474

AC Q20474

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)

DE F46F6.3 PROTEIN.

GN F46F6.3.

OS CAENORHABDITIS ELEGANS.

OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;

OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.

OC [1]

RP SEQUENCE FROM N.A.

RX COTTAGE A.;

RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

RL [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL NATURE 368:32-38(1994).

DR EMBL: Z50028; E1346872; .

SQ SEQUENCE 158 AA; 18019 MW; C3A480F2 CRC32;

Query Match 70.5%; Score 43; DB 5; Length 158;

Best Local Similarity 55.6%; Pred. No. 1.68e+01;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 83 QDPPESEP 91

||||:|

QY 1 QENPDSSEP 9

RESULT 10

ID Q24743

AC Q24743

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE EGP-1 PRECURSOR.

GN EGP-1.

OS DROSOPHILA VIRILIS (FRUIT FLY).

OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;

OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;

OC DROSOPHILIDAE; DROSOPHILA.

OC [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;

RA THUREROFF E., STOEVEN S., KRESS H.;

RL MECH. DEV. 37:81-93(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VIRILIS; TISSUE=Salivary Gland;  
 RX MEDLINE; 90276249.  
 RA SWIDA U., LUCKA L., KRESS H.;  
 RT "Glue protein genes in Drosophila virilis: their organization,  
 development and control of transcription and specific mRNA  
 degradation.";  
 RL DEVELOPMENT 108:269-280(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VIRILIS; TISSUE=Salivary Gland;  
 RX MEDLINE; 90384577.  
 RA KRESS H., SWIDA U.;  
 RT "Drosophila glue protein gene expression. A proposal for its  
 ecysone-dependent developmental control.";  
 RL NATURWISSENSCHAFTEN 77:317-324(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VIRILIS; TISSUE=Salivary Gland;  
 RA JARRIN HENTSCHEL A., THUEROFF E., TISCHENDORF B., KRESS H.;  
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VIRILIS; TISSUE=Salivary Gland;  
 RX MEDLINE; 94200049.  
 RA KRESS H.;  
 RT "The salivary gland chromosomes of Drosophila virilis: a cytological  
 map, pattern of transcription and aspects of chromosome evolution.";  
 RL CHROMOSOMA 102:734-742(1993).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VIRILIS; TISSUE=Salivary Gland;  
 RX MEDLINE; 90276248.  
 RA KRESS H., LUCKA L., SWIDA U., THUEROFF E., KLEMM U.;  
 RT "Genes from two intermolt puffs in Drosophila virilis polytene  
 chromosomes are differentially transcribed during larval  
 development.";  
 RL DEVELOPMENT 108:261-267(1990).  
 DR EMBL; 249942; G887426;  
 DR FLYBASE; FBgn0005594; Dvir\egp1.  
 KW SIGNAL.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT SIGNAL 1 18 POTENTIAL.  
 SQ SEQUENCE 184 AA; 20567 MW; C0E0E0FB CRC32;  
 Query Match 70.5%; Score 43; DB 5; Length 184;  
 Best Local Similarity 50.0%; Pred. No. 1.68e+01;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Db 77 BESPEDEPVP 86  
 QY 1 QENPDSSEPV 10  
 RESULT 11  
 ID O88346 PRELIMINARY; PRT; 262 AA.  
 AC O88346;  
 DT 01-NOV-1998 (TREMREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE SLOW SKELETAL MUSCLE TROPONIN T.  
 GN TNNT1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129SVJ; TISSUE=SKELETAL MUSCLE;  
 RX MEDLINE; 98322260.  
 RA JIN J.-P., CHEN A., HUANG Q.-Q.;  
 RT "Three alternatively spliced mouse slow skeletal muscle troponin T

RT isoforms: conserved primary structure and regulated expression during  
 postnatal development";  
 RL GENE 214:121-129(1998).  
 DR EMBL; AF020946; G3449362;  
 SQ SEQUENCE 262 AA; 31344 MW; 392A6A01 CRC32;  
 Query Match 70.5%; Score 43; DB 11; Length 262;  
 Best Local Similarity 50.0%; Pred. No. 1.68e+01;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Db 23 EEAPPEPVP 32  
 QY 1 QENPDSSEPV 10  
 RESULT 12  
 ID Q27033 PRELIMINARY; PRT; 480 AA.  
 AC Q27033;  
 DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE MEMBRANE PROTEIN.  
 OS THEILERIA PARVA.  
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; PIROPLASMA; THEILERIIDAE;  
 OC THEILERIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MUGUGA;  
 RX MEDLINE; 94088665.  
 RA BAYLIS H.A., ALLSOPP B.A., HALL R., CARRINGTON M.;  
 RT "Characterisation of a glutamine- and proline-rich protein (QP  
 protein) from Theileria parva";  
 RL MOL. BIOCHEM. PARASITOL. 61:171-178(1993).  
 DR EMBL; L06323; G310893;  
 KW MEMBRANE.  
 SQ SEQUENCE 480 AA; 52388 MW; 1EDC7723 CRC32;  
 Query Match 70.5%; Score 43; DB 5; Length 480;  
 Best Local Similarity 40.0%; Pred. No. 1.68e+01;  
 Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
 Db 97 QOGPTPOPI 106  
 QY 1 QENPDSSEPV 10  
 RESULT 13  
 ID O88573 PRELIMINARY; PRT; 1217 AA.  
 AC O88573;  
 DT 01-NOV-1998 (TREMREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE PROTO-ONCOGENE AFG.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M; TISSUE=THYMUS;  
 RA ISNARD P.;  
 RT "Mouse AFG coding sequence";  
 RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AF074266; G3328190;  
 SQ SEQUENCE 1217 AA; 131774 MW; 15E914D6 CRC32;  
 Query Match 70.5%; Score 43; DB 11; Length 1217;  
 Best Local Similarity 50.0%; Pred. No. 1.68e+01;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 541 QOHPSKDPL 550  
 QY 1 QENPDSSEPV 10



```
RESULT 14
ID P97579 PRELIMINARY; PRT; 1382 AA.
AC P97579;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HEPATOCYTE GROWTH FACTOR RECEPTOR.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY;
RA LIU Y., TOLBERT E.M., SUN A.M., DWORKIN L.D.;
RL AM. J. PHYSIOL. 271:0-0(1996).
RL EMBL: U65007; G1679660;
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 1382 AA; 153750 MW; 500939CA CRC32;

Query Match 70.5%; Score 43; DB 11; Length 1382;
Best Local Similarity 60.0%; Pred. No. 1.68e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 349 QSKPDSAEPM 358
QY 1 QENPDSSEPV 10

RESULT 15
ID P97523 PRELIMINARY; PRT; 1382 AA.
AC P97523;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HGF RECEPTOR PRECURSOR.
GN C-MET.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER;
RX MEDLINE; 97419268.
RA WALLENIS V., RAWET H., SKRTIC S., EKBERG S., HELOU K., QIU Y.,
RA LEVAN G., CARLSSON B., ISAKSSON O., NAKAMURA T., JANSSEN J.O.;
RT "Chromosomal localization of rat hepatocyte growth factor (Hgf) and
RT HGF receptor (Met) and characterization of Hgf receptor cDNA.";
RL MAMM. GENOME 8:661-667(1997).
DR EMBL; X96786; E238809;
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM; PF00069; pkinase; 1.
KW SIGNAL.
FT SIGNAL. 1 24 POTENTIAL.
FT CHAIN 25 1382 HGF RECEPTOR.
SQ SEQUENCE 1382 AA; 153940 MW; ED5D6941 CRC32;

Query Match 70.5%; Score 43; DB 11; Length 1382;
Best Local Similarity 60.0%; Pred. No. 1.68e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 349 QSKPDSAEPM 358
QY 1 QENPDSSEPV 10
```

Search completed: Thu Oct 21 15:48:53 1999  
Job time : 50 secs.

=> s aspartyl beta-hydroxylase?

FILE 'USPAT'

2322 ASPARTYL  
182906 BETA  
1198 HYDROXYLASE?  
L2 0 ASPARTYL BETA-HYDROXYLASE?  
(ASPARTYL (W) BETA (W) HYDROXYLASE?)

FILE 'USOCR'

47 ASPARTYL  
4398 BETA  
30 HYDROXYLASE?  
L3 0 ASPARTYL BETA-HYDROXYLASE?  
(ASPARTYL (W) BETA (W) HYDROXYLASE?)

FILE 'JPO'

303 ASPARTYL  
32326 BETA  
78 HYDROXYLASE?  
L4 0 ASPARTYL BETA-HYDROXYLASE?  
(ASPARTYL (W) BETA (W) HYDROXYLASE?)

FILE 'EPO'

324 ASPARTYL  
19279 BETA  
223 HYDROXYLASE?  
L5 0 ASPARTYL BETA-HYDROXYLASE?  
(ASPARTYL (W) BETA (W) HYDROXYLASE?)

TOTAL FOR ALL FILES

L6 0 ASPARTYL BETA-HYDROXYLASE?